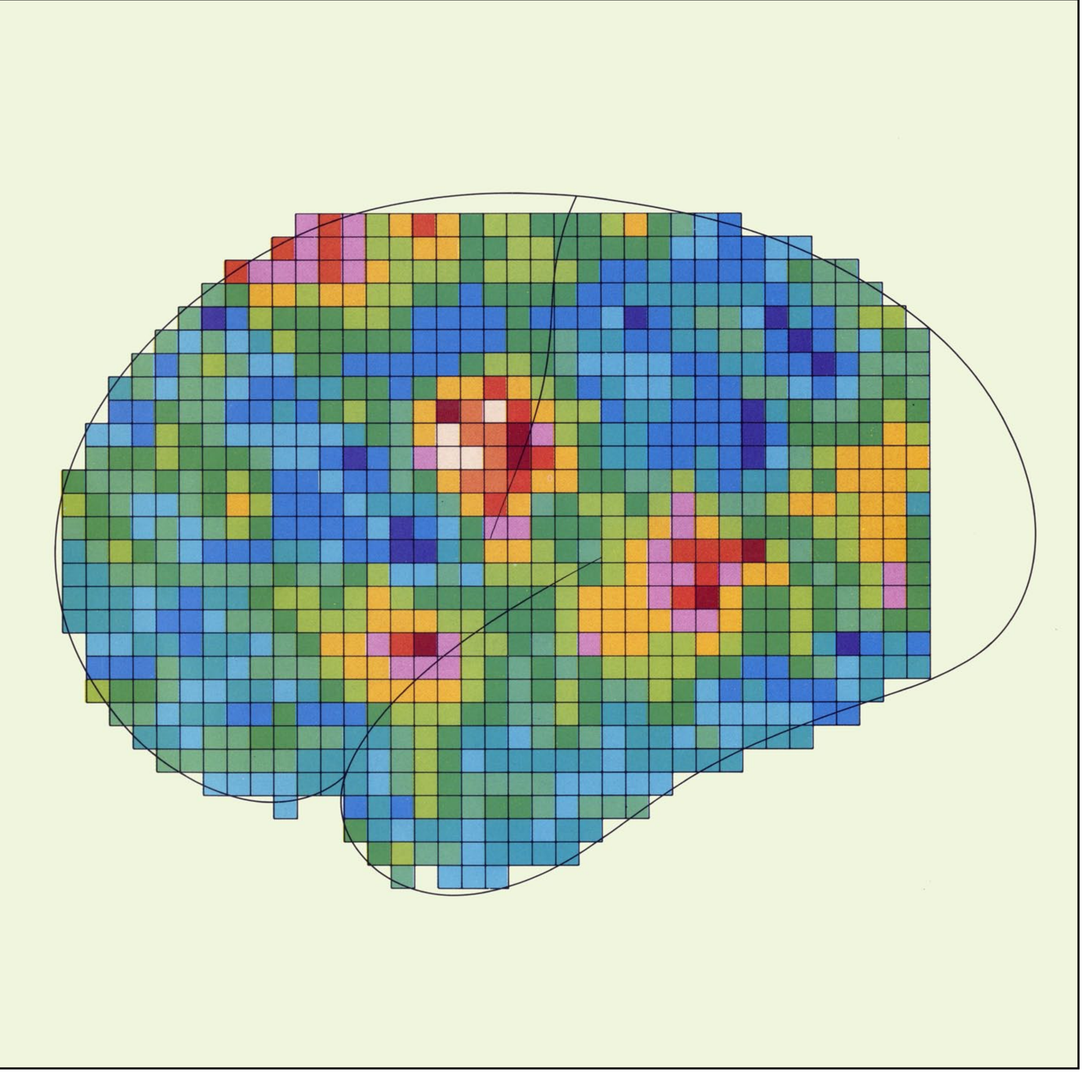
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



FUNCION CEREBRAL Y FLUJO SANGUINEO

Diciembre 1978

170 PTAS.

Copyright © 1978 Prensa Científica S.A.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 27
Diciembre 1978

INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN

SCIENTIFIC AMERICAN

COMITE DE REDACCION: Gerard Piel (Presidente); Dennis Flanagan, Francis Bello, Philip Morrison; Trudy E. Bell; Judith Friedman; Brian P. Hayes; Jonathan B. Piel; John Purcell; James T. Rogers; Armand Schwab, Jr.; Jonathan B. Tucker; Joseph Wisnovsky

DIRECCION EDITORIAL: Dennis Flanagan
DIRECCION ARTISTICA: Samuel L. Howard
PRODUCCION: Richard Sasso
DIRECTOR GENERAL: Donald H. Miller, Jr.

INVESTIGACION Y CIENCIA

DIRECTOR: Francisco Gracia Guillén
REDACCION: José María Valderas Gallardo (Redactor Jefe)
Encarna de Blas (Secretaria de Redacción)
PRODUCCION: César Redondo Zayas
PROMOCION Y PUBLICIDAD: Elena Sánchez-Fabrés
PROMOCION EXTERIOR: Pedro Clotas Cierco
EDITA: Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

Luis Palacios: Función cerebral y flujo sanguíneo; Pedro Pascual: *La partícula hypsilón;* Luis Bou: *El último teorema de Fermat;* Manuel Puigcerver: *¿Desencadenó una supernova la formación del sistema solar?* Antonio Blanco: *Los dibujos prehistóricos en el suelo del Perú;* Ramón Fontarnau y Frances E. Lynd: *La formación de modelos en el desarrollo biológico;* Joandomènec Ros: *Una migración en masa de aves terrestres sobre el océano;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio.*

Hace...

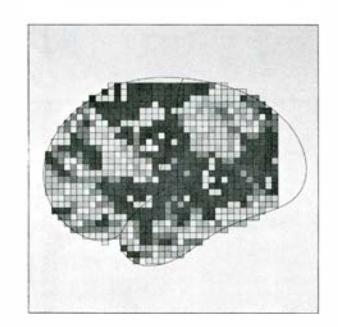
J. M. López Piñero

Ciencia y sociedad:

Ramón Bayés, Mercedes Durfort y Eduardo Cadenas

Libros:

Antonio Alabau, Enrique Julve y Manuel Fernández



LA PORTADA

El dibujo de la portada reproduce una imagen generada por ordenador que representa el modelo de actividad del lado izquierdo de la corteza cerebral de un ser humano normal durante la ejecución de un tipo de comportamiento específico (en este caso, leer en voz alta). Dado que el volumen de riego sanguíneo en una zona particular de la corteza es proporcional a la actividad de las células nerviosas en dicha zona, las funciones de la corteza se pueden localizar inyectando el isótopo radiactivo xenón 133 en una arteria del cerebro y midiendo el riego regional de sangre en la corteza durante la ejecución de tareas sensoras, motoras o mentales mediante una serie de detectores de centelleo (véase "Función cerebral y flujo sanguíneo", por Niels A. Lassen, David H. Ingvar y Erik Skinhøj en este mismo número). Cada pixel –cuadrado de la figura– representa un centímetro cuadrado de corteza. Los colores simbolizan el nivel de riego sanguíneo; en esta reproducción de la figura el blanco y el rojo indican activación intensa y, el azul obscuro, fuerte inhibición. La imagen revela que el leer en alta voz requiere la colaboración de siete zonas corticales discretas, cada una especializada.

Suscripciones:
Prensa Científica, S.A.
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 1870 pesetas
Extranjero:
Un año (12 números): 30 U.S.$
Ejemplar atrasado ordinario:
190 pesetas
Ejemplar atrasado extraordinario:
270 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49; Barcelona-10

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239; Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovín
Avda. de Moratalaz, 137; Madrid-30
Tel. 439 41 50 ext. 468
Cataluña:
Miguel Munill
Balmes, 191, 2.º, 3.ª; Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Ilustración de la portada de Gabor Kiss

Página	Fuente
7-15	Pedro Pastor Lozano, Eduardo Lastra L. Arenosa y Antonio Sellés
19	Niels A. Lassen
20-22	Carol Donner
23	Medimatic A/S
24	Medimatic A/S (*arriba*) Carol Donner (*abajo*)
25-26	Niels A. Lassen
27	Carol Donner
31	Fermi National Accelerator Laboratory
32-33	Allen Beechel
34	Ben Rose (*arriba*), Allen Beechel (*abajo*)
35-36	Allen Beechel
37	Fermi National Accelerator Laboratory (*arriba*), Allen Beechel (*abajo*)
38-39	Allen Beechel
45	Dominique Berretty, Black Star
46	Butler Library
47-48	New York Public Library
50	Springer-Verlag
55	David N. Schramm
56-60	George V. Kelvin
61	George V. Kelvin (*arriba*) William Herbst y George Assousa (*abajo*)
62-64	George V. Kelvin
69	Servicio Aerofotográfico Nacional
70-71	Alan D. Iselin
72	Servicio Aerofotográfico Nacional (*abajo*), Wilfredo Loayza (*abajo*)
73	William H. Isbell
74-75	Servicio Aerofotográfico Nacional (*arriba*), Alan D. Iselin (*abajo*)
79	Lewis Wolpert
80-82	Alan D. Iselin
83	Alan D. Iselin (*arriba*) Lewis Wolpert (*abajo*)
84	Alan D. Iselin
85	Alan D. Iselin (*arriba*) Lewis Wolpert (*abajo*)
86	Alan D. Iselin (*arriba, izquierda y abajo*), Lewis Wolpert (*arriba, derecha y centro*)
87	Alan D. Iselin
91	Timothy C. Williams y Janet M. Williams
92-93	Lorelle M. Raboni
95-98	Lorelle M. Raboni
105	Jearl Walker
106-110	Michael Goodman

Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona.
Printed in Spain - Impreso en España

Copyright © 1978 Scientific American, Inc., 415 Madison Av., New York, N.Y. 10017.

Copyright © 1978 Prensa Científica, S.A., Calabria, 235-239 - Barcelona-29 (España).

Reservados todos los derechos. Prohibida la reproducción en todo o en parte por ningún medio mecánico, fotográfico o electrónico, así como cualquier clase de copia, reproducción, registro o transmisión para uso público o privado, sin la previa autorización escrita del editor de la revista

El nombre y la marca comerciales SCIENTIFIC AMERICAN, así como el logotipo distintivo correspondiente, son propiedad exclusiva de Scientific American, Inc., con cuya licencia se utilizan aquí.

Los autores

EDUARDO LASTRA L. ARENOSA y PEDRO PASTOR LOZANO ("Transmisión digital de señales") trabajan desde hace varios años en el Centro de Investigación y Estudios de la Compañía Telefónica Nacional de España, como subdirector y jefe de equipos de transmisión y fuerza, respectivamente. Ambos son ingenieros superiores de telecomunicación por la Escuela Técnica Superior de Ingenieros de Telecomunicación de Madrid, y cuenta con amplia experiencia en el campo de la transmisión telefónica en general, y particularmente en el estudio y desarrollo de equipos de transmisión digitales.

NIELS A. LASSEN, DAVID H. INGVAR y ERIK SKINHØJ ("Función cerebral y flujo sanguíneo") han colaborado en investigaciones de circulación sanguínea cerebral en el hombre durante 20 años. Lassen es jefe del departamento de fisiología clínica del Hospital Bispebjerg en Copenhague. Obtuvo sendos doctorados en medicina y en ciencias por la Universidad de Copenhague; en 1957 estudió la circulación cerebral con Seymour S. Kety, que estaba entonces en el National Institute of Mental Health de los Estados Unidos. Ingvar es profesor y jefe del departamento de neurofisiología clínica en la Universidad de Lund en Suecia. Se graduó en medicina y en ciencias en Lund, y desde 1951 hasta 1953 fue colaborador en el Instituto Neurológico de Montreal bajo la dirección de Wilder Penfield y Herbert Jasper. Skinhøj es profesor y director del departamento de neurología de la Universidad de Copenhague. Siendo estudiante, durante la Segunda Guerra Mundial, tomó parte en el movimiento de resistencia y estuvo preso en un campo de concentración alemán. Después de la guerra volvió a la Universidad de Copenhague, en la que se doctoró. Los autores desean expresar su agradecimiento a sus muchos colaboradores, en particular a Edda Sveinsdottir, de la Universidad de Copenhague.

LEON M. LEDERMAN ("La partícula hypsilón") es "Higgins Professor of Physics" en la Universidad de Columbia y director del acelerador Nevis en Columbia. Hizo sus estudios de licenciatura en el City College de Nueva York y obtuvo su doctorado en Columbia en 1951. Lederman ha participado en la obtención del primer haz de piones en un acelerador, en el descubrimiento de la partícula *K* neutra, en la observación de la violación de la paridad en las desintegraciones del pion y del muon, en la medida del momento magnético anómalo del muon, en el descubrimiento de una segunda clase de neutrinos y en el hallazgo de las primeras indicaciones de hadrones con gran momento transversal.

HAROLD M. EDWARDS ("El último teorema de Fermat") enseña matemáticas en la Universidad de Nueva York. Comenzó sus estudios en la Universidad de Wisconsin, licenciándose por la de Columbia en 1957. Se doctoró en Harvard en 1961. Tras cuatro años de docencia en Columbia, pasó en 1966 a la universidad neoyorquina. Los principales campos de investigación de Edwards son la teoría de números y la historia de las matemáticas. Es autor de *Advanced Calculus* (1969), *Riemann's Zeta Function* (1974) y *Fermat's Last Theorem* (1977). Edwards desea agradecer a la Fundación Vaughn y la National Science Foundation su ayuda económica para realizar la investigación tema de su artículo.

DAVID N. SCHRAMM y ROBERT N. CLAYTON (¿Desencadenó una supernova la formación del sistema solar?) forman parte ambos del profesorado de la Universidad de Chicago. Schramm es catedrático y director del departamento de astronomía y astrofísica. Recibió su grado de Bachelor del Instituto de Tecnología de Massachusetts y su doctorado del Instituto de Tecnología de California. Tras dos años en la Universidad de Texas en Austin se trasladó a Chicago en 1974. Los intereses de investigación de Schramm incluyen el origen de los elementos químicos en explosiones de supernova, evolución estelar y galáctica, las consecuencias de colapso gravitatorio y el enlace entre la cosmología y la física de los núcleos atómicos y las partículas de alta energía. Clayton es catedrático y director del departamento de ciencias geofísicas y también profesor de química. Nacido en el Canadá, realizó sus estudios universitarios elementales en la Universidad de la Reina (Queen's University) y obtuvo su doctorado en química en el Instituto de Tecnología de California (Cal Tech). Clayton pasó entonces dos años como profesor de la Universidad del Estado de Pennsylvania antes de entrar a formar parte del profesorado de la de Chicago en 1958.

WILLIAM H. ISBELL ("Los dibujos prehistóricos en el suelo del Perú") es profesor asistente de antropología en la Universidad de Nueva York, en Binghamton. Se graduó en la Universidad estatal de San Francisco y recibió su grado de doctor en filosofía de la Universidad de Illinois, en 1973. Desde 1966 ha investigado los geoglifos, pirámides en plataformas y varios tipos de arquitectura civil del Perú en un esfuerzo para entender "los cambios que tuvieron lugar en las instituciones sociales que regulaban la población y el consumo de energía cuando los gobiernos prehistóricos se hicieron más centralizadores y jerárquicos." En el momento actual, William Isbell dirige el Proyecto de Prehistoria Urbana de Huari, participación estadounidense en el proyecto conjunto, a largo plazo, peruano-norteamericano para la investigación y conservación de Huari, la capital urbana del primer gran estado del Perú.

LEWIS WOLPERT ("La formación de modelos en el desarrollo biológico") enseña biología aplicada en la Middlesex Hospital Medical School de Londres. Nacido en Sudáfrica, estudió ingeniería en la Universidad de Witwatersrand. Trabajó luego en el National Building Research Institute, Gran Bretaña, durante dos años y pasó un año en el Water Planning Departmet de Israel. No llegándole a satisfacer la ingeniería, preparó su tesis doctoral sobre la mecánica de la división celular con J. F. Danielli en el Kings College de Londres. Se integró luego en la facultad de zoología del Kings College y trabajó sobre el movimiento ameboide y la regeneración de la hidra. Wolpert pasó a la Middlesex Hospital Medical School en 1966.

TIMOTHY C. WILLIAMS y JANET M. WILLIAMS ("Una migración en masa de aves terrestres sobre el océano") constituye un equipo de investigación formado por marido y mujer. Timothy Williams es profesor asociado de biología en el Swarthmore College. Graduado por Swarthmore, obtuvo su grado de licenciatura por la Universidad de Harvard y su doctorado por la Universidad de Rockefeller en 1968. Después de pasar siete años enseñando en la Universidad estatal de Nueva York en Buffalo y un año trabajando como científico visitante en la Woods Hole Oceanographic Institution, Williams se incorporó en 1976 a la facultad de Swarthmore. Janet Williams es investigadora independiente en el Marine Biological Laboratory en Woods Hole. Se graduó en el Colby College y obtuvo su grado de licenciatura por la Universidad de Nueva York en 1967.

Hace...

...cuatrocientos años

Juan López de Velasco, cosmógrafo del Consejo de Indias, publicó su *Instrucción y advertimiento para la observación de los eclipses de la Luna, y cantidad de las sombras, que S. M. mandó hacer este año de 1578 en las ciudades y pueblos españoles de las Indias, para verificar la longitud y altura dellos.*

Dicha instrucción formaba parte de uno de los programas científicos más ambiciosos de la época, consagrado a la investigación sistemática de la geografía americana. Se proponía, por un lado, reunir una detallada descripción de la geografía física y humana, mediante "relaciones" de los distintos territorios minuciosamente planificadas. Pretendía, por otro, verificar la longitud y la latitud del mayor número posible de poblaciones. Para ello, López de Velasco redactó "instrucciones" en 1577, 1578 y 1584, poniendo especial cuidado de explicar minuciosamente procedimientos de observación astronómica asequibles a personas de escasa o nula preparación científica. De esta forma, fue posible reunir gran cantidad de datos sobre localidades dispersas en territorios de una extensión inédita en este tipo de estudios. En la ejecución del programa participaron numerosas figuras científicas, entre ellas, el gran ingeniero Juanelo Turriano y el tratadista de náutica Rodrigo Zamorano, catedrático en la Casa de Contratación de Sevilla. No obstante, la importancia de la empresa residió en que se recogieron y sistematizaron series muy amplias de observaciones realizadas con métodos muy sencillos.

...trescientos años

Apareció la *Practica Gotholanorum* (1678) de Jacint Andreu, catedrático de Prima de Medicina en la Universidad de Barcelona. Su labor docente, durante casi un cuarto de siglo, había contribuido notablemente a la elevación del nivel científico de la medicina catalana, a pesar de que no tenía en absoluto una mentalidad renovadora. Ni siquiera admitía el uso de los medicamentos químicos. "¡Ojalá esta pestilente y epidémica plaga de químicos –afirma– fuera desterrada de toda España!" Sin embargo, no se dedicó a refutar las doctrinas modernas, como hicieron otros galenistas ortodoxos, sino que centró todo su esfuerzo en la observación clínica. Gracias a ello, su libro es un excelente tratado de medicina práctica con descripciones clínicas claras, concisas y, en buena parte, basadas en su experiencia personal. Entre ellas, figura la que generalmente se considera como la primera exposición detallada del tétanos de los recién nacidos, que Andreu llama "barretas".

...doscientos años

Inició sus trabajos la Expedición Botánica al Perú que, bajo la dirección de Hipólito Ruiz y de José Antonio Pavón, estudió la flora peruana y chilena desde 1778 a 1.788. Fue una de las grandes expediciones científicas españolas de esta época, comparable, entre las centradas en la botánica, a la que José Celestino Mutis dirigió en Nueva Granada y a la que Martín de Sessé encabezó en México.

Ruiz, Pavón y sus colaboradores recogieron cerca de tres mil especímenes vegetales y realizaron millares de dibujos y descripciones. Dieron a conocer al mundo científico varios centenares de géneros y especies nuevos, muchos de los cuales conservan en la actualidad los nombres que ellos les dieron. Los resultados los expusieron en numerosos trabajos, a partir de una polémica *Quinología*, aparecida en 1792. De la publicación oficial de la expedición, titulada *Flora Peruviana, et Chilensis*, apareció primero un *Prodromus* (1794) con la presentación de 136 géneros nuevos y, más tarde, tres volúmenes con materiales peruanos (1798-1802).

...cien años

Francisco Quiroga Martínez obtuvo los títulos de licenciado en ciencias naturales y de doctor en farmacia. Discípulo de José Macpherson, era ya entonces una figura destacada de la geología y, en especial, de la mineralogía. Sobresalió en la investigación microscópica de la estructura cristalina de los minerales, tema sobre el que publicó una excelente monografía cuando solamente tenía veintidós años (1875). Realizó importantes estudios sobre ofitas, basaltos, limburgitas, etcétera, que le proporcionaron un notable prestigio internacional. Ello fue la base de que se creara en la Universidad de Madrid la primera cátedra de cristalografía que hubo en Europa, que Quiroga ocupó hasta su temprana muerte en 1894, a los cuarenta y un años de edad.

...cincuenta años

Pío del Río-Hortega fue nombrado jefe de la Sección de Investigación del Instituto Nacional del Cáncer de Madrid, del que pasaría a ser director tres años más tarde. Su principal tema de trabajo en la citada institución fue la histopatología de los tumores generados en el sistema nervioso. Le consagró una serie de publicaciones, que aparecieron principalmente en la revista *Archivos Españoles de Oncología*, fundada por el propio Río-Hortega y editada por el Instituto. Su contribución más importante fue quizá la tipología de los gliomas y paragliomas de acuerdo con la distinta madurez de sus elementos celulares. Aunque renovador, este capítulo de la obra de Río-Hortega no tuvo tanta repercusión como sus estudios en torno a la neuroglía, quizá por su desconexión de la clínica.

Transmisión digital de señales

El advenimiento de las técnicas de transmisión digitales ha supuesto una auténtica revolución en las telecomunicaciones. Desde la invención de la válvula termoiónica, sólo se habían registrado avances en la transmisión analógica

Pedro Pastor Lozano y Eduardo Lastra L. Arenosa

Tradicionalmente, las señales procedentes de una fuente de información analógica (es decir, aquella cuyos valores varían de forma continua con el tiempo, como la voz) han sido transmitidas también en forma analógica. Desde hace algunos años, la aplicación de las técnicas digitales a la transmisión de esas señales ha revolucionado el mundo de las telecomunicaciones. Mediante tales técnicas, las señales analógicas se transforman, por conversión, en señales digitales (en funciones que adoptan sólo valores discretos con el tiempo), representadas por dígitos o bits, tal como la información procedente de señales telegráficas o de datos. La unidad de información selectiva se llama bit. Es la cantidad de información que se deriva del conocimiento de la aparición de uno de dos sucesos equiprobables, exclusivos y exhaustivos. Esto permite una utilización más eficaz de los diferentes medios de transmisión y un tratamiento homogéneo de las señales con independencia de cuál sea la fuente de información original (voz, vídeo, telegrafía o datos). En este artículo se presentan las técnicas digitales de transmisión basadas en la modulación por impulsos codificados (MIC), partiendo de su justificación a la luz de la teoría de la información, y exponiendo su desarrollo y su incidencia en las telecomunicaciones.

Consideremos una comunicación en la forma conceptual más simple. Por una parte, se tendrá el origen o fuente de la información; por otro lado, el destino de la misma, y, entre ambos, el canal de comunicación encargado de enlazarlos entre sí.

Dentro del canal de comunicación hay que distinguir tres partes básicas. La primera, que denominaremos genéricamente transmisor, encargada de modificar o cambiar la naturaleza de las señales que recibe de la fuente. La segunda, el medio de transmisión, que recibe las señales convenientemente modificadas por el transmisor y las transporta hasta el punto deseado. Y por último el receptor, que recoge las señales del medio, las vuelve a transformar en su primitiva naturaleza y las entrega en su destino. En todos estos procesos que tienen lugar en el canal de comunicación, las señales son perturbadas por diversas distorsiones y por la aparición de otras señales espúreas indeseables. Esto requiere decir que, en el extremo receptor, la señal aparecerá mezclada con otras componentes surgidas durante la transmisión, y ajenas, por tanto, a la información original. Todos estos efectos se conocen por el nombre genérico de ruido. El grado de calidad de la comunicación, por lo que al ruido se refiere, dependerá de la diferencia que exista, en el extremo de destino, entre la potencia de la señal y la de dicho ruido, que la enmascara.

Para hacer que la información llegue con la suficiente calidad, dos son las actuaciones posibles. Una es conseguir que el ruido acumulado durante la comunicación no resulte excesivo. Dicho en términos más correctos, conseguir que la relación entre la potencia de la señal y la potencia del ruido no sea inferior a un determinado valor límite. Otro recurso es el de aumentar la redundancia de la información que se transmite. La redundancia puede definirse como un exceso de información, sobre la estrictamente necesaria, para caracterizar un mensaje. Por ejemplo, es corriente cuando se habla por teléfono y se quiere transmitir una palabra de entendimiento difícil (un nombre o apellido poco usual) el deletrearla, asociando a cada letra un nombre conocido o preestablecido del que aquélla es inicial: "... *B* de Barcelona, *E* de España, *M* de Madrid... etcétera". Esto es una forma elemental de redundancia, ya que, para transmitir una sola palabra, utilizamos una información mucho más copiosa que la imprescindible. En este caso, la redundancia la aplica directamente la persona que habla, pero con el mismo principio, el canal de comunicación puede incluir, entre los procesos a que somete las señales, una forma de operación más o menos redundante, que permita compensar el efecto de pérdida de información producido por el ruido.

Volviendo al esquema básico de una comunicación, hemos dicho que el transmisor modificaba la naturaleza de la información antes de entregarla al medio de transmisión. ¿Por qué es necesaria esta transformación? Hay dos razones fundamentales para ello. La primera, y más importante desde el punto de vista teórico del proceso de comunicación, es la necesidad de adecuar la naturaleza de las señales a transmitir a la del medio de transmisión. En el caso de la telefonía, por ejemplo, las señales eléctricas directamente obtenidas de la fuente están situadas dentro de la banda de frecuencias que corresponde a la voz humana. Pero es evidente que este margen de frecuencias puede no ser el más indicado, o incluso puede resultar absolutamente inadecuado, para ser transmitido por un determinado medio. Es preciso, por tanto, modificar tales señales para hacerlas corresponder al margen de trabajo de dicho medio. La segunda razón, de índole económica, la constituye la conveniencia de aprovechar al máximo un determinado medio de transmisión, a base de transmitir por él, no sólo una comunicación, sino varias. Se trata, pues, de empaquetar varias comunicaciones, y transmitir simultáneamente el conjunto. Para realizar esta operación, denominada multiplexación, es preciso, como en el caso anterior, transformar las señales.

Resumiendo, las funciones básicas que realiza el transmisor en un canal de comunicación son: adecuar la naturaleza de las señales al medio empleado, llevar a cabo la multiplexación necesaria entre varias comunicaciones y proporcionar la "redundancia" precisa, según el sistema de comunicación de que se trate. Para desarrollar estas operaciones se utilizan diversas técnicas de modulación. En un proceso de modulación, se debe trasladar la información que lleva una determinada señal (moduladora) a otra señal (portadora) que es la que reúne las características que se desean.

Por ello, se actúa sobre algunos de los parámetros de esta segunda, al objeto de que varíe en función de la información que contiene la primera.

El principal parámetro que caracteriza un medio de transmisión es el constituido por la banda de frecuencia que es capaz de transmitir en debidas condiciones. Si consideramos una información procedente de una fuente analógica (voz, música, etcétera), resulta que estará formada por un cierto conjunto de frecuencias comprendidas en una banda bastante baja (inferior a 15 kilohertzs). Si se desea transmitir por un medio de transmisión cuya banda de trabajo sea de, por ejemplo, 300 a 500 kilohertzs, habrá que conseguir que esa información se imprima sobre un conjunto de frecuencias que caiga dentro de esa banda.

Se trata, por tanto, de conseguir que una señal analógica se transforme en otra, también analógica, pero de frecuencia muy diferente. En este caso, de transformación de una señal analógica en otra analógica, tres son las técnicas básicas que se nos ofrecen: la modulación de amplitud, de frecuencia y de fase. Mediante la primera, se toma una señal (portadora) y se hace que su amplitud varíe en función de la señal inicial (moduladora). Mediante la segunda, se hace variar la frecuencia de la portadora (dentro de un determinado entorno), en función de la señal moduladora. En el tercer caso, lo que se altera en la portadora en función de la moduladora es la fase.

Si la señal inicial en lugar de analógica es impulsiva (la información está representada entonces por una sucesión de impulsos), podremos transferir esta información a una portadora de una forma similar, haciendo variar su amplitud o su frecuencia. La única diferencia con lo anterior estará en que esa variación de amplitud o de frecuencia de la portadora, en lugar de tener carácter analógico (es decir, de poder adoptar cualquier valor dentro del margen de excursión), tendrá carácter discreto, adoptando sólo un número determinado de valores. Este es el caso de la llamada telegrafía armónica, en la que los estados de presencia o ausencia de una tensión continua se transforman en presencia o ausencia de una portadora (modulación de amplitud), o en presencia de frecuencia alta y presencia de frecuencia baja (modulación de frecuencia).

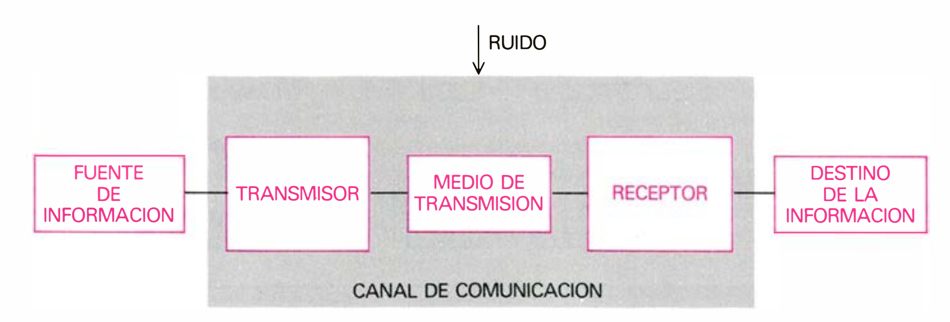
Supongamos, finalmente, que la señal inicial es analógica, y se quiere transformarla en impulsiva. En este caso, deberemos utilizar una portadora de carácter impulsivo, que está formada por una serie de impulsos iguales, separados entre sí en un determinado intervalo de tiempo. Así como en una portadora analógica o sinusoidal son tres los parámetros que la caracterizan (amplitud, frecuencia y fase), que dan lugar a los tres correspondientes tipos de modulación, en una señal impulsiva cabe considerar otros tres: la amplitud de los impulsos, la anchura de los mismos y la posición que ocupa, o puede ocupar, cada uno respecto a la teórica. En el caso de hacer variar la amplitud de los impulsos en función de la señal moduladora, se obtiene la modulación por impulsos en amplitud. Si es la anchura de cada impulso la que adopta un valor proporcional a la amplitud de la moduladora en ese instante, se obtiene la modulación por impulsos en anchura. Finalmente, puede hacerse que la posición de cada impulso varíe, respecto a la teórica, en función de la moduladora. Así se obtiene la modulación de impulsos por posición.

Existe una cuarta forma de modulación impulsiva, cuya filosofía es algo más elaborada y compleja que la de los tres precedentes. Partiendo de una señal impulsiva modulada en amplitud, se procede a asignar a cada impulso un valor numérico en función de la amplitud que presenta. Seguidamente, se representa ese número mediante una palabra de código binario, a base de un conjunto de impulsos de amplitud fija, por ausencia o presencia de los mismos ("unos" y "ceros", es decir, bits o dígitos). Este procedimiento, denominado modulación por impulsos codificados (MIC, o PCM en la literatura inglesa), es el que ha encontrado un extraordinario eco en el mundo de las telecomunicaciones, por las razones que posteriormente se verán.

He aquí, expuesto en forma muy sucinta, el repertorio de criterios básicos de modulación que es posible emplear para la transformación de una señal analógica en otra analógica, de una impulsiva en otra analógica, o de una señal analógica en otra impulsiva. Mediante la aplicación de alguno de estos criterios, o de combinaciones de ellos, puede transformarse cualquier señal de manera que pueda ser aceptada y transmitida por cualquier medio de transmisión. Pero hay otra razón fundamental para transformar una señal, además de la de adecuarla a la naturaleza del medio transmisor. Esta razón es la de transmitir varias comunicaciones simultáneamente por el mismo canal de comunicación. Este proceso se denomina multiplexación. Es fácil entender cómo puede conseguirse utilizando los conceptos de modulación vistos.

Mediante uno o varios procesos de modulación se puede traspasar la información original a una señal situada en cualquier margen de frecuencias prefijado. Así pues, si las señales correspondientes a una serie de comunicaciones (telefónicas, por ejemplo, y situadas todas en el mismo margen de frecuencias), se someten a un proceso de modulación, cada una con una frecuencia portadora diferente, cada información quedará ubicada en un banda, también diferente, de frecuencias. No habrá, entonces, inconveniente alguno en transmitirlas conjuntamente por el mismo medio, ya que al estar en bandas distintas no pueden mezclarse o confundirse. En el extremo opuesto, bastará separar cada porción de frecuencias para recuperar, individualizada, la información correspondiente a cada comunicación. Esta es, expresada en forma muy sumaria, la idea elemental de la multiplexación por distribución de frecuencia.

Pero también podemos utilizar una técnica de modulación que conduzca a una señal impulsiva, en lugar de analógica. En ese caso, la multiplexación puede concebirse de otra manera. Puesto que entre dos impulsos consecutivos existe un determinado espacio de tiempo, pueden ubicarse en él impulsos correspondientes al proceso de modulación de otras comunicaciones. De esa forma, el primer impulso corresponderá a la primera comunicación, el segundo a la segunda, y así sucesivamente, hasta que el enésimo más uno vuelva a corresponder a la primera, el enésimo más



ESQUEMA DE UNA COMUNICACION en su forma conceptual más sencilla. Por una parte, se tiene el origen de la información, por otra el destino de la misma, y, entre ambos, el canal de comunicación encargado de enlazarlos entre sí. El transmisor se encarga de modificar o cambiar la naturaleza de las señales que recibe de la fuente. El medio de transmisión recibe las señales convenientemente modificadas por el transmisor y las transporta hasta el punto deseado. Por último, el receptor recoge las señales del medio, las vuelve a transformar en su primitiva naturaleza y las entrega en su destino. Estos procesos se dan en el canal de comunicación.

dos a la segunda, etcétera. De esta manera se obtiene el concepto básico de la multiplexación por distribución en el tiempo.

Imaginemos un sistema ideal de comunicación, a través del cual queremos transmitir una cantidad de información I. Los principales parámetros a considerar aquí son: la anchura de la banda de frecuencias que el sistema es capaz de transmitir y que llamaremos B; el tiempo T, necesario para realizar la transmisión de la cantidad de información citada; y la relación P/R entre la potencia de la señal transmitida y la del ruido. La expresión que relaciona todos estos parámetros es:

$$I = T \, . \, B \, . \, \log_b(1 + P/R)$$

Esta fórmula fue establecida por Shannon, en 1948, y determina el comportamiento de un canal ideal de comunicación. La elección de la base de logaritmos b origina tres tipos de unidad de información: BIT (Binary unit) para b - 2; NAT (Natural unit) para b - e y DECIT (Decimal unit) aún llamado HARTLEY para b - 10. El bit es la unidad más utilizada. De acuerdo con ella vemos, por ejemplo, que si fijamos un determinado valor para la relación señal-ruido (P/R), puede transmitirse la misma cantidad de información I por un canal de ancho de banda inferior en un tiempo mayor. De la misma forma, un canal que lleve asociado poco ruido no necesita tener un ancho de banda tan dilatado como otro cuyo ruido sea mayor.

La fórmula de Shannon antes expuesta puede transformarse en la siguiente:

$$C = B \log_b(1 + P/R)$$

donde $C = I/T$ designa la capacidad de información del canal, es decir, la cantidad de información por unidad de tiempo que puede transmitir. Esta expresión regula el comportamiento de un canal de transmisión ideal. Pero, ¿qué ocurre en un canal real? Cualquier sistema, dependiendo fundamentalmente de la técnica de modulación que utilice, presentará un comportamiento más o menos aproximado al ideal. Tomemos unos ejes coordenados cartesianos, de forma que en el eje de abscisas figure la relación P/R entre la potencia de la señal y la del ruido, y en el de ordenadas, el cociente C/B, es decir, la capacidad de información por unidad de ancho de banda. Representemos en dicho sistema el comportamiento del canal ideal, según la fórmula de Shannon.

Si en el mismo gráfico representamos el comportamiento (obtenido empíricamente) correspondiente a canales que utilicen diversas técnicas de modulación, aparecerá un conjunto de curvas del tipo de las que se ven en la ilustracióm de esta página. En esa representación puede apreciarse claramente que la técnica de modulación MIC es la que más se aproxima a la curva correspondiente al canal ideal. Podría desarrollarse todo un estudio analítico matemático comparativo, a través del cual apareciesen, de un modo objetivo, las características diferenciales de las diversas técnicas de modulación. Pero únicamente se ha pretendido mostrar (ya que no demostrar) de qué manera puede llegarse a la conclusión de que la técnica MIC presenta ventajas conceptuales superiores a otras técnicas de modulación, como preámbulo a la exposición detallada que de tal técnica vamos a desarrollar.

La modulación por impulsos codificados (MIC) es una técnica relativamente reciente en cuanto a su utilización generalizada. No obstante, ya en el si-

TECNICAS BASICAS DE MODULACION. En un proceso de modulación, se debe trasladar la información que lleva una determinada señal (moduladora) a otra señal (portadora) que es la que reúne las características que se desean. Por ello, se actúa sobre algunos de los parámetros de esta segunda, al objeto de que varíe en función de la información que contiene la primera. La ilustración muestra gráficamente los tres tipos de transformación: de una señal analógica a otra también analógica (*arriba*), de una señal impulsiva a otra señal analógica (*en el medio*) y de una señal analógica a otra impulsiva (*abajo*). En los dos primeros casos se utilizan técnicas de modulación mediante portadora analógica y, en el tercero, mediante portadora impulsiva.

glo XIX, y basándose en procedimientos de tipo telegráfico, se llevaron a cabo experiencias encaminadas a intentar transmitir voz y música mediante señales eléctricas codificadas. Con independencia de estos lejanos antecedentes, que no fructificaron en absoluto debido al rudimentario estado de la técnica electrónica en aquella época, fue en el año 1937 cuando A. H. Reeves desarrolló la técnica de modulación por impulsos codificados. Ni siquiera entonces estaba lo suficiente avanzada la tecnología de los componentes electrónicos (tubos termoiónicos) como para favorecer la extensión de esta nueva técnica. Hasta que no se generalizó el uso de componentes de estado sólido, ya en la segunda mitad del presente siglo, la técnica MIC no pudo abrirse paso, con un éxito creciente y arrollador.

Veamos las diversas etapas seguidas en el proceso MIC. Partimos de la señal a transmitir, procedente de una conversación telefónica, por ejemplo, y constituida por una onda de baja frecuencia de amplitud variable de forma continua. A esta señal es preciso someterla primero a un proceso de muestreo. Es obvio que no podemos representar por impulsos el valor de la señal en todos los instantes, puesto que obtendríamos un espectro continuo, confundiéndose unos impulsos con otros. Debemos elegir ciertos puntos de esa señal separados por un cierto intervalo de tiempo, y representar la amplitud que en ese momento tenga mediante un impulso cuya amplitud sea proporcional a ella.

Importa subrayar que si el muestreo se realiza de forma correcta, las muestras no sólo representan una buena aproximación de la señal original, sino que contienen absolutamente toda la información de esa señal. Esto queda reflejado en un teorema de la teoría de la información, denominado teorema de muestreo, y que dice: "Si una información, que es una magnitud función continua del tiempo, se muestrea instantáneamente a intervalos regulares a una frecuencia que sea al menos dos veces la frecuencia más alta de dicha información, las muestras obtenidas contienen toda la información original".

Aunque, de forma muy sucinta, explicaremos el alcance de dicho principio. Puesto que cualquier función del tiempo puede considerarse descompuesta en una serie de funciones sinusoidales, imaginemos que tratamos de muestrear una sinusoide de frecuencia *F*. Consideremos un período *T* de ella. Si de esa porción de la curva fijamos dos puntos, la curva queda totalmente definida. O dicho de otra manera, sólo hay dos valores como máximo que sean independientes y, fijados éstos, todos los demás quedan determinados. Entonces, si la curva en cuestión queda definida por dos puntos cualesquiera, bastará con transmitir dos muestras para transmitir la información completa de la sinusoide. El muestrear dos puntos por cada período *T* equivale a muestrear con una frecuencia doble de la frecuencia *F*. Si este concepto lo aplicamos a la frecuencia más alta a transmitir, del conjunto de ellas en que puede descomponerse la señal original será tanto más válido para frecuencias más bajas, obteniéndose así la conclusión enunciada por el teorema de referencia. Esta primera etapa de muestreo es común a todas las técnicas de modulación por impulsos. Las etapas siguientes, en cambio, son específicas de la técnica MIC.

El paso siguiente es la operación de cuantificación. El conjunto de impulsos modulados en amplitud, que se obtiene como consecuencia del muestreo, presenta la característica de que las amplitudes de tales impulsos varían de forma analógica, o lo que es lo mismo, pueden adoptar cualquier valor. Esto representa un importante inconveniente a la hora de asignar una palabra código que represente cada una de esas amplitudes, ya que se precisaría un sistema de codificación con infinitos valores posibles. Hay que conseguir que sólo exista un número discreto o finito de posibles valores, y para ello debemos cuantificar tales amplitudes. Es decir, se trata de definir una escala de valores fijos y de asignar a cada amplitud, que estará comprendida entre dos valores cuánticos sucesivos, aquel de los dos que se halle más cerca de su valor real. Esta operación introduce una inevitable distorsión (o ruido), tanto mayor cuanto más elevado sea el intervalo entre valores cuánticos contiguos. Esta distorsión, no obstante, está perfectamente controlada, puesto que tendrá un valor constante, función de la escala de cuantificación que se elija, y bastará entonces con mantenerlo por debajo del límite que se estime conveniente, y que puede ser tan pequeño como se desee.

La tercera operación básica, una vez realizada la cuantificación, es la codificación. Consiste esta etapa en sustituir cada muestra por un grupo de impulsos de amplitud fija, que por la combinación de presencias o ausencias de los mismos forma una determinada palabra código (de carácter binario), que representa de modo inequívoco el valor cuantificado de dicha muestra. Se obtiene así, por último, un tren de impulsos de amplitud fija, es decir, cuya amplitud no contiene ninguna información, ya que ésta reside en la presencia o ausencia de cada impulso. Este conjunto de impulsos se envía al medio de transmisión, como depositario, en forma codificada, de la información original.

Las tres operaciones descritas (muestreo, cuantificación y codificación) constituyen las etapas básicas necesarias para el proceso de modulación MIC. Sin embargo, existen otras dos operaciones de carácter complementario (en cuanto a que no serían imprescindibles para la filosofía del proceso), pero de importancia significativa, ya que completan y perfeccionan los resultados obtenidos por las otras, aumentando sustancialmente la eficacia global. La primera de ellas está relacionada con la cuantificación, y tiende a compensar el efecto siguiente. Ocurre que la cantidad de ruido introducida por el proceso de cuantificación es constante y depende del número de niveles cuánticos elegido, o lo que es lo mismo, de la magnitud del intervalo entre dos niveles sucesivos, pero no de la amplitud de la señal. Por tanto, en las amplitudes de valor bajo, la relación señal/ruido será mucho peor que en las amplitudes de mayor magnitud, ya que el denominador es siempre el mismo. Este distinto valor de la relación señal/ruido de unas muestras respecto a otras es totalmente indeseable.

Si se fija un número de escalones cuánticos bastante elevado para que la relación señal/ruido en niveles bajos de señal sea aceptable, estaremos obteniendo una relación señal/ruido innecesariamente buena en los niveles altos, y pagando el precio de complicar la codificación, ya que aumentamos el número de palabras código necesarias. La solución óptima aparece clara. Es necesario realizar, asociado a la cuantificación, un proceso que asigne intervalos cuánticos más separados a las amplitudes altas y, más juntos, a las bajas. Es decir, que, en vez de hacer una asignación proporcional o lineal de niveles en función del valor de amplitud, esta asignación se realice según una ley no lineal. A este proceso se le conoce por compresión, ya que su efecto consiste en comprimir los intervalos cuánticos que corresponden a valores bajos de la amplitud de las muestras.

Por último, la segunda operación complementaria, que se realiza al final del proceso, es decir, después de la etapa de codificación, tiene por objeto alterar la polaridad de los impulsos de salida. Veamos por qué. Tal como se ha descrito el proceso de modulación MIC hasta ahora, a la salida del codificador se obtiene un tren de impulsos unipolares, es decir, todos ellos con la misma polaridad. Esto presenta el inconveniente de que una señal así formada contiene un gran porcentaje de su energía en componentes de baja frecuencia, cuya transmisión presenta en general serios inconvenientes. Para resolver dicha dificultad, basta someter a estos impulsos a una operación simple consistente en invertir alternadamente su polaridad, de

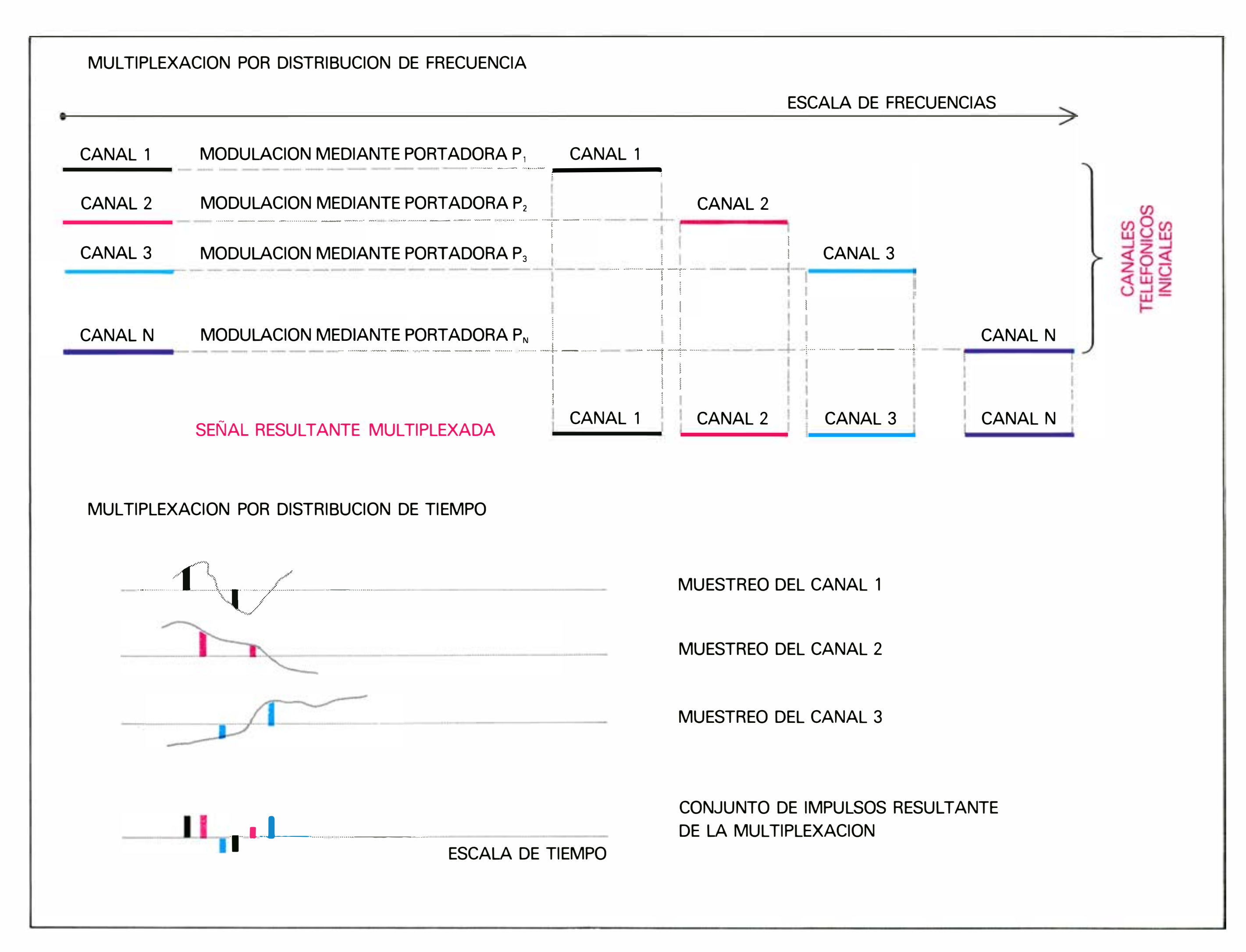
manera que dos impulsos consecutivos tengan siempre polaridad opuesta. El conjunto resultante, de naturaleza bipolar, ya no contiene esos componentes indeseables.

Queda con esto completa la exposición, hecha en la forma conceptual más simple, de todo el proceso de modulación MIC, desde la señal original en baja frecuencia hasta el tren de impulsos codificados portadores de la información inicial, a través de las etapas de muestreo, cuantificación (con la comprensión asociada), codificación y alternancia de polaridad. Dado que el proceso incorpora la representación de la información por valores discretos, posteriormente representados por una palabra de código binario, es decir, al fin y a la postre por números, se ha extendido la denominación de señales numéricas o señales digitales para las procedentes de un proceso MIC y la de sistemas numéricos o digitales para aquellos sistemas de transmisión basados en dicha técnica.

La utilización de la modulación MIC, como la de todas las técnicas de modulación, está justificada por un doble motivo: por trasvasar la información a señales de naturaleza más adecuada o ventajosa para el medio de transmisión que se va a utilizar, y para aplicar el proceso de multiplexación que permita la transmisión simultánea de varias comunicaciones por un mismo portador. Ya vimos que el principio básico de la multiplexación por distribución de tiempo consiste en intercalar, en los espacios de tiempo comprendidos entre dos impulsos consecutivos de un tren de ellos correspondiente a una comunicación, los impulsos correspondientes a otras comunicaciones.

En un sistema de transmisión digital pueden siempre distinguirse dos partes diferentes: el equipo multicanal encargado de formar el paquete que contenga la combinación de las señales correspondientes a un determinado número de comunicaciones, y el equipo de línea, que transporta ese paquete o señal compleja de un punto a otro, para ser demodulado en el equipo multicanal distante, a fin de individualizar así las comunicaciones originales. Por razones variadas de flexibilidad, economía, eficacia y otras, el proceso de multiplexación se realiza escalonadamente. En una primera etapa, se forma un conjunto con un determinado número de comunicaciones. Si este número es suficiente para cubrir la capacidad deseada, el proceso ha terminado, y la señal digital compuesta por los impulsos combinados resultantes de todas esas comunicaciones puede entregarse al equipo de línea para su transmisión. Si, por el contrario, se desea una capacidad mayor, se realiza un nuevo proceso de multiplexación; sólo que ahora, en lugar de multiplexar las señales analógicas correspondientes a comunicaciones simples, se multiplexan conjuntos de impulsos como los obtenidos en la etapa anterior. Procediendo sucesivamente así, pueden irse obteniendo señales digitales complejas que contengan la información de un número creciente de comunicaciones.

La ley de formación de los sucesivos órdenes de multiplexación responde a normalizaciones de carácter internacional, con vistas a facilitar las conexiones entre diversos países y la compatibilidad entre equipos procedentes de distintos fabricantes.



LAS DOS TECNICAS DE MULTIPLEXACION posibles son las de "distribución de frecuencia" y "distribución de tiempo". Mediante la primera, se trasladan los distintos canales a bandas de frecuencia diferentes. Mediante la segunda, "por distribución de tiempo", se intercalan entre sí las muestras correspondientes a los distintos canales. El proceso de multiplexación de las muestras permite la transmisión simultánea de varias comunicaciones por un mismo portador.

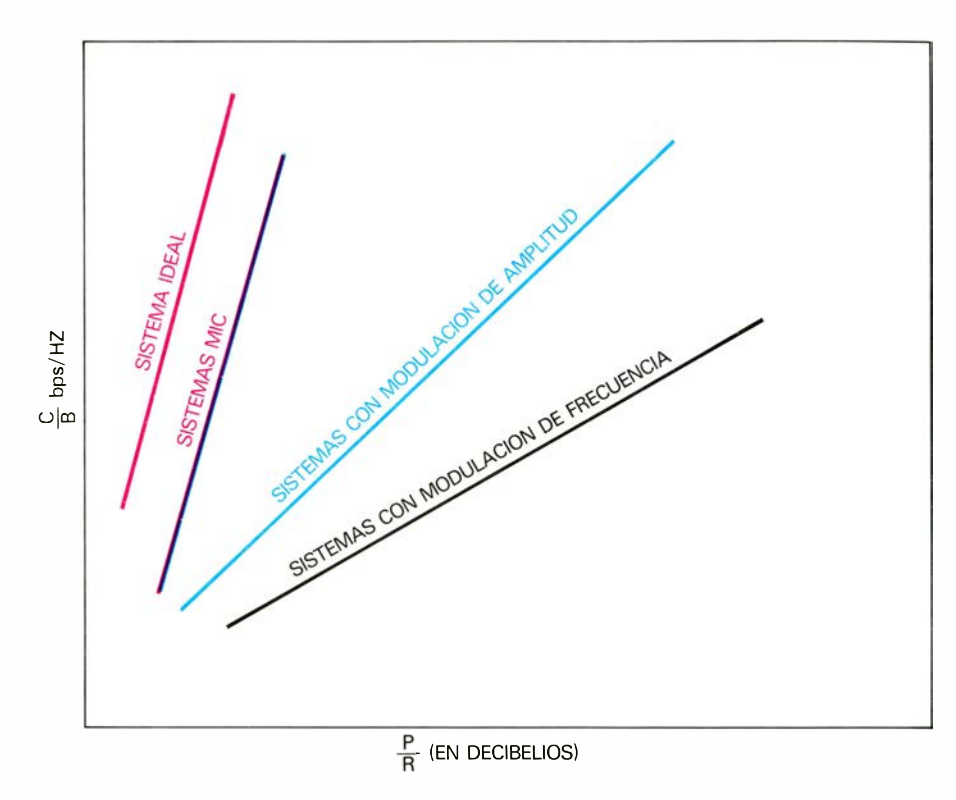
Veamos, en forma resumida, las características principales de esta normalización. En el terreno de las telecomunicaciones, dos son los organismos internacionales que emiten recomendaciones o normativas, aceptadas en Europa: el Comité Consultivo Internacional Telegráfico y Telefónico (CCITT), dependiente de la Unión Internacional de Telecomunicaciones, y la Conferencia Europea de Correos y Telecomunicaciones (CEPT). Las recomendaciones de esta última suelen ser compatibles con las de la primera, pero de forma que las del CCITT son más amplias, o con más grados de libertad, que las de la CEPT.

Veamos el primer escalón de multiplexación MIC (primario). El CCITT recomienda, como número de canales a multiplexar, 24 o 30. El sistema de 24 canales es el elegido por la ATT (American Telephone and Telegraph, empresa norteamericana de explotación telefónica), y se utiliza en Estados Unidos y Japón. El sistema de 30 canales es el normalizado por la CEPT; se usa en Europa.

Ambos sistemas emplean, como frecuencia de muestreo, la de 8 kilohertzs (doble del ancho de banda del canal telefónico), muestreándose, por tanto, cada comunicación cada 125 microsegundos. En este intervalo de tiempo, que se denomina trama, se introducen las muestras de las demás comunicaciones. En el sistema europeo, la trama se divide en 32 intervalos de tiempo de 3,9 microsegundos cada uno, de los que 30 contienen los impulsos codificados de las muestras correspondientes a otros tantos canales telefónicos, y los dos restantes contienen impulsos de sincronismo y señalización. En el sistema americano, la trama se divide en 24 intervalos de tiempo, cada uno de 5,18 microsegundos, correspondientes a otros tantos canales telefónicos junto con su señalización, y otro intervalo de 0,65 microsegundos para sincronización.

Cada muestra se representa mediante ocho impulsos en forma codificada binaria (por ausencia o presencia de ellos). Como consecuencia, la velocidad de transmisión (número de impulsos por unidad de tiempo) de la señal digital compleja, a la salida del proceso de multiplexación, es de 2048 kilobits por segundo en el sistema europeo y de 1544 kilobits por segundo en el americano.

El proceso de compresión asociado a la cuantificación se realiza según una curva de transferencia, que en el caso europeo responde a una ley de tipo logarítmico denominada "ley A", y en el americano a otra, también logarítmica pero diferente, denominada "ley μ". En ambos casos, la cuantificación se realiza con 256 niveles, que es el número máximo posible de palabras código diferentes utilizando ocho impulsos de codificación ($2^8 = 256$).



REPRESENTANDO LA VARIACION de la capacidad de información por unidad de ancho de banda ***(C/B)*** **de un canal, en función de la relación señal ruido** ***P/R*** **en el mismo, puede verse que la técnica MIC es la que más se aproxima al comportamiento del canal ideal.** ***C*** **indica la capacidad de un canal (cantidad de información por unidad de tiempo);** ***B*****, ancho de banda;** ***P*****, potencia de la señal y,** ***R*****, potencia del ruido.**

La codificación suele efectuarse después de la multiplexación de las muestras, aunque se puede codificar también las muestras de cada canal antes de la multiplexación con codificadores individuales por cada canal. Esto último tiene la ventaja de permitir mayor flexibilidad en la constitución de la señal digital. Es posible, por ejemplo, sustituir un canal telefónico normal (analógico) y en su lugar insertar las señales de un canal digital de información, cuyo régimen de impulsos es de 64 kilobits por segundo (es decir, 8 kilohertzs multiplicados por 8 bits). Actualmente están apareciendo codificadores individuales económicos, gracias a las posibilidades que ofrece la tecnología integrada monolítica en el campo de los componentes de estado sólido. No obstante, su uso no está aún muy extendido.

Para finalizar con la etapa primaria de multiplexación, queda por mencionar un detalle importante. Cualquier sistema multicanal por distribución en el tiempo requiere un procedimiento que garantice que el transmisor de un extremo y el receptor del otro trabajen de forma sincronizada. Con tal propósito, se inserta en cada trama, de forma periódica (de período igual al de la trama o múltiplo de éste), un intervalo de tiempo que contenga una palabra código característica que permita identificar cada canal. En el sistema europeo se utilizan ocho bits que se transmiten en una trama sí y en otra no, con una frecuencia de repetición de 4 kilohertzs. Además, para sincronizar la señalización telefónica de cada canal, se insertan cuatro bits cada 16 tramas, con una frecuencia de repetición de 500 Hz.

Veamos ahora las etapas siguientes de multiplexación (secundaria, terciaria, etcétera). En la etapa primaria que se acaba de describir, los equipos múltiplex (recientemente se ha extendido la denominación "muldex", ya que incluyen *mul*tiplexación y *de*multiple*xa*ción), convierten, incluyendo un proceso MIC, varias señales analógicas en una señal digital, y viceversa. Sin embargo, en las etapas sucesivas de multiplexación, los muldex ya no recogen señales analógicas, sino que combinan varias señales digitales (provenientes de la etapa anterior), mediante multiplexación por distribución en el tiempo, en una única señal digital de mayor velocidad. A estos muldex, que no contienen proceso MIC, se los denomina digitales a diferencia de los primarios, reconocidos por muldex MIC.

En los muldex digitales, la multiplexación se lleva a cabo mediante entrelazado de los impulsos. Es decir, las diversas señales digitales originales (señales tributarias) se combinan bit a bit. Esta solución es más eficaz que el entrelazado de intervalos de tiempo entre sí, que precisaría de un almacenamiento previo de las señales.

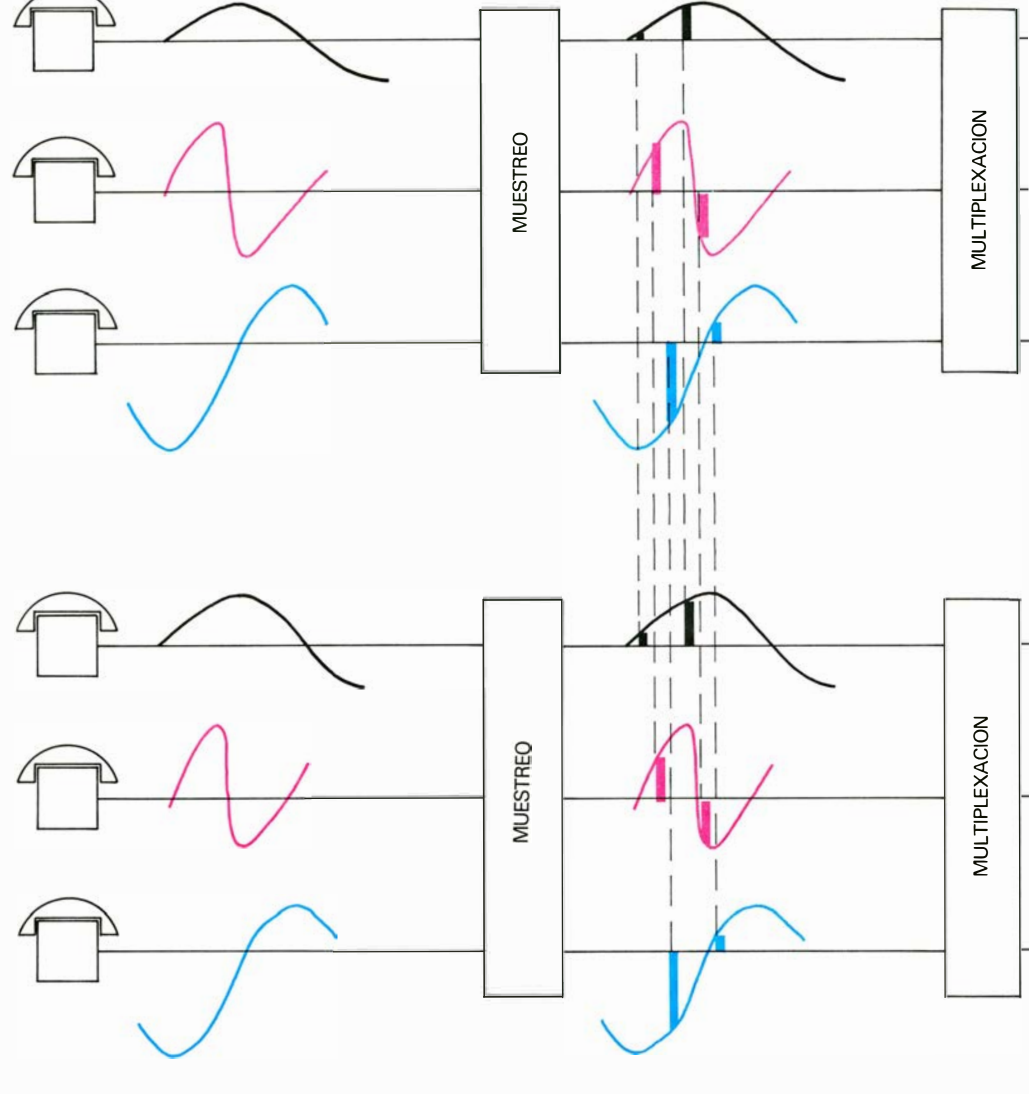
También en la multiplexación de órdenes superiores al primario hay que tener en cuenta el problema del sincronismo. Sucede que, aunque las fuentes originales sean síncronas, las señales no lo son del todo a causa de los retrasos causados por la transmisión. Deben, por tanto, considerarse como plesiócronas (es decir, que tienen la misma cadencia nominal, pero con unos márgenes de variación específicos; del griego *plesio*, que significa casi).

El procedimiento para multiplexar estas señales plesiócronas consiste en crear una trama formada por el entrelazado de los bits de cada señal, más una información auxiliar de relleno que permita sincronizarlos. En el receptor, la demultiplexación se realiza con ayuda de una palabra código de sincronismo como en las etapas primarias, restituyendo las señales originales y suprimiendo el relleno introducido. En el argot técnico, se suele utilizar, para designar ese relleno, el término justificación, palabra extraída del contexto de las artes gráficas y que hace referencia al proceso de ajuste de los espacios entre letras para que todos los renglones tengan la misma longitud.

En definitiva, la velocidad de la señal digital de salida (V_S) de una etapa de multiplexación por división en el tiempo vendrá dada por la expresión:

$$V_S = n V_T + K$$

en donde V_T es la velocidad digital de las señales tributarias de entrada, n el número de ellas y K el número de bits por segundo necesarios para servicio (incluyendo tanto los de relleno o justificación como los de sincronismo).



SISTEMA DE MODULACION DE IMPULSOS CODIFICADOS (MIC). En la figura se representa un sistema de transmisión telefónico basado en la técnica de modulación de impulsos codificados. Las etapas de

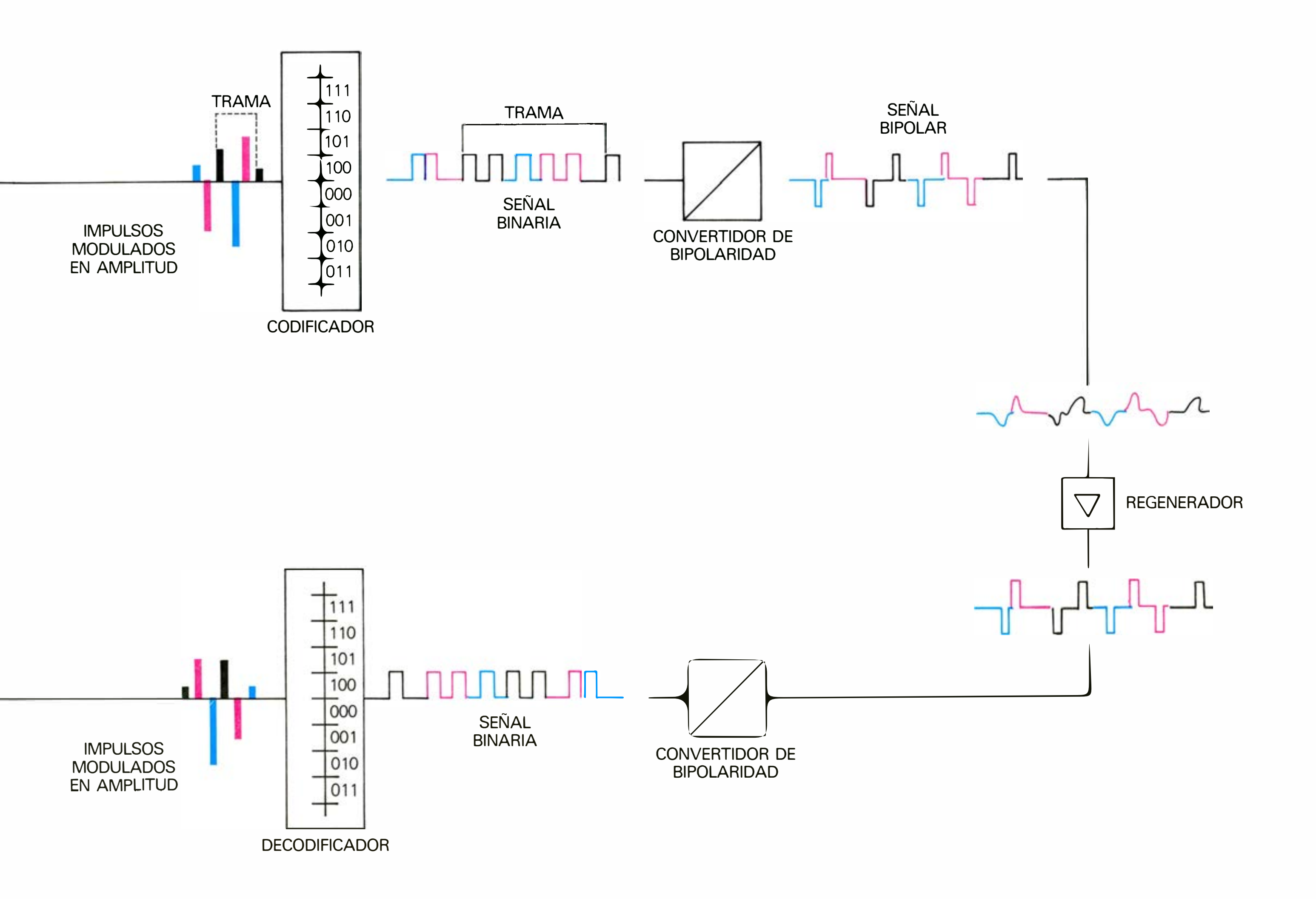
La organización de los distintos niveles jerárquicos de multiplexación digital arranca de la expresión citada, tomando como punto inicial de partida la velocidad de la etapa primaria (1544 kilobits por segundo o 2048 kilobits por segundo según se parta del sistema americano o del europeo). A la hora de definir las velocidades de las diversas etapas, se han suscitado fuertes debates en el seno de los organismos internacionales, debates que aún hoy en día continúan, y que han originado una cierta dispersión en las jerarquías normalizadas hasta la fecha. Por lo que a Europa se refiere, la jerarquía MIC se basa en el primario de 2048 kilobits por segundo (30 canales), y parte de un criterio de uniformidad en cuanto al factor multiplicativo del número de canales de un orden al del siguiente, que es siempre el cuatro. En consecuencia, hoy existen los cuatro órdenes jerárquicos normalizados siguientes:

(*1*) Primario, multiplexación MIC de 30 canales telefónicos, con una velocidad de la señal digital saliente, de 2048 kilobits por segundo. (*2*) Secundario, multiplexación digital de cuatro señales tributarias primarias, es decir, 120 canales telefónicos, con una velocidad de la señal digital saliente de 8448 kilobits por segundo. (*3*) Terciario, multiplexación digital de cuatro señales tributarias secundarias, es decir, 480 canales con una velocidad de la señal digital saliente de 34.368 kilobits por segundo. (*4*) Cuaternario, multiplexación digital de cuatro señales tributarias terciarias, es decir, 1920 canales telefónicos, con una velocidad de la señal digital saliente de 139.264 kilobits por segundo. La única excepción europea a esta norma la constituye Francia, que elimina el escalón terciario, pasando directamente de 120 canales a 1920 canales.

Paralelamente a esta jerarquización, existen muldex MIC de 8448 kilobits por segundo, es decir, que transforman directamente 120 canales telefónicos, en la señal digital secundaria, y estructuras de trama MIC de primero y segundo orden (2048 Kb/s por segundo, o 8448 Kb/s, respectivamente) formadas por 32 o 132 canales digitales de 64 kilobits por segundo, para centrales de conmutación temporal, tanto para telefonía como para señales de datos. En todos estos niveles, se utiliza relleno o justificación positiva (que es cuando la velocidad de la señal multiplexada es mayor que la nominal).

Aparte de esta normalización europea, existen otras que toman como base el primario de 24 canales. Los niveles normalizados hasta la fecha son: (*1*) Primario: 24 canales telefónicos, velocidad de 1544 Kb/s. (*2*) Secundario: 96 canales telefónicos, velocidad de 6312 Kb/s. (*3*) Terciario: en este nivel hay dos versiones: la norteamericana, 672 canales telefónicos, velocidad de 44.736 Kb/s., y la japonesa, 480 canales telefónicos, velocidad 32.004 Kb/s. Finalmente, la Unión Soviética piensa en una jerarquía idéntica a la europea en cuanto a velocidades, pero con un procedimiento de relleno o justificación diferente.

Se ha hablado, hasta aquí, de la parte multicanal de los sistemas. Veamos ahora algunas consideraciones sobre la otra parte, el equipo de línea. Una vez que el equipo multicanal ha terminado sus ope-



muestreo, multiplexación, codificación-decodificación y conversión de bipolaridad se han realizado sólo con tres canales telefónicos y considerando tres bits (bit es la unidad de información selectiva) por muestra, ocho niveles cuánticos (2^3). La cuantificación es lineal con objeto de clarificar el proceso.

raciones de multiplexación, se obtiene una señal digital compleja, combinación de todos los impulsos correspondientes a los distintos canales telefónicos, más los de sincronismo, y otros. Esta señal tendrá una velocidad mayor o menor según las etapas de multiplexación que se hayan llevado a cabo. En todo caso, velocidad aparte, su naturaleza es siempre la misma: un conjunto de impulsos de amplitud constante. El equipo de línea se encarga de recoger esta señal, transportarla hasta el punto de destino y entregarla allí, en buenas condiciones, al equipo múltiplex receptor, para su demultiplexación.

El equipo de línea consta de dos tipos de elementos: los repetidores y el portador. Por este último habrá de circular la señal digital a lo largo del trayecto deseado, y dado que durante el camino dicha señal se irá debilitando, será preciso intercalar periódicamente repetidores que compensen tal debilitamiento. Si la señal fuese analógica, la acción de dichos elementos se reduciría a amplificarla, aumentando su potencia. Esto tiene el efecto de que todas las imperfecciones que la señal lleve asociadas al llegar a un repetidor (ruido, distorsiones, etcétera), son también amplificadas junto con la señal, y, por tanto, son acumulativas. La actuación a seguir con una señal digital es muy diferente. En ella, la información no reside en la amplitud o forma de los impulsos, sino en la presencia o ausencia de éstos. Por tanto, por muy imperfectos que éstos lleguen al repetidor, con tal de que aún sea identificable su presencia, es suficiente. El repetidor entonces, al constatar la llegada de un impulso, genera otro nuevo, y lo transmite. Es decir, vuelve a generar realmente la señal original; es como si el transmisor lo hubiésemos trasladado a ese punto; y las perturbaciones que pudieran acompañar a la señal desaparecen. Por esta razón, se denomina regeneradores a los repetidores MIC.

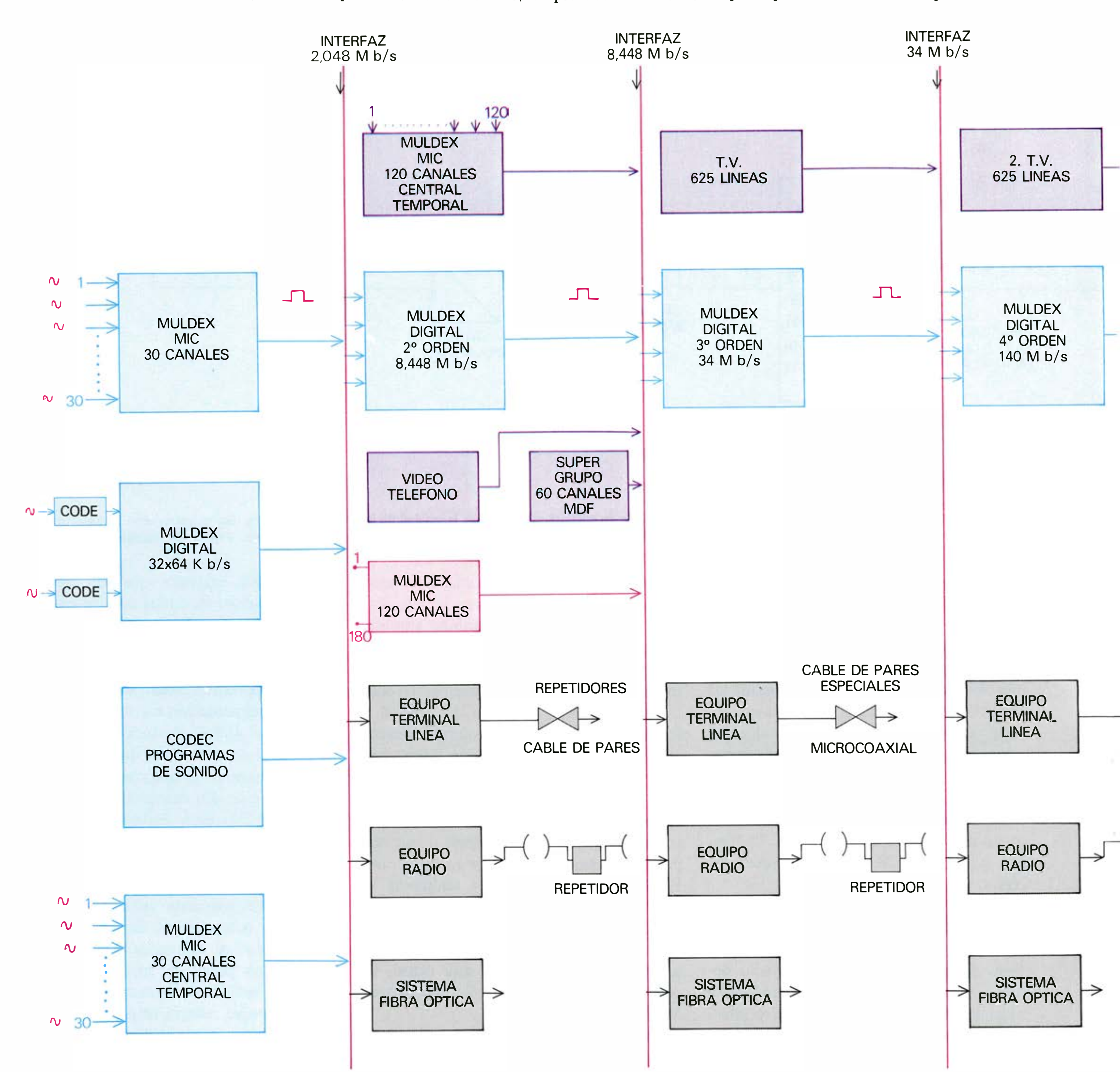
El conjunto de regeneradores distribuidos a lo largo de la ruta, y que realizan la función que acaba de exponerse (para ambos sentidos de transmisión), se completa con los regeneradores terminales, uno en cada extremo, que son equipos más complejos, asociados con los equipos múltiplex, y que, además de dicha función de regeneración, efectúan algunas operaciones complementarias, aunque no menos importantes, como son la de suministrar la alimentación a los regeneradores intermedios (telealimentación) y proveer un sistema de telecontrol que permita detectar las averías de tales regeneradores y su localización. En cuanto al portador, las señales MIC pueden utilizar en principio cualquier medio convencional (cable de pares coaxial, radio, etcétera).

Los primeros sistemas MIC encontraron su más adecuada aplicación en aumentar la potencialidad de los cables de pares simétricos existentes. Los sistemas utilizan dos pares de conductores, uno para cada sentido de transmisión. Debido a la inmunidad relativa respecto al ruido e interferencias, puede utilizarse cualquier tipo de cable. Unicamente habrá que tener en cuenta las características del cable utilizado a fin de elegir la distancia máxima a la que hay que colocar los regeneradores y

el número máximo de sistemas que pueden coexistir en pares distintos del mismo cable. En cables de buena calidad, se puede llegar a ocupar hasta el 40 por ciento de sus pares con sistemas MIC. En cables de baja calidad, esta ocupación máxima puede bajar hasta un 15 por ciento.

Más tarde, se empezaron a utilizar, cuando era preciso instalar cable nuevo, unos tipos especialmente diseñados para sistemas MIC, y en los cuales la mitad de los pares del cable está separada de la otra mitad mediante una pantalla electrostática (de aluminio o similar), de forma que su comportamiento equivale al de dos cables separados. Con un grupo de pares para un sentido de transmisión, y el otro para el opuesto, la ocupación de sistemas MIC puede alcanzar casi al 100 por ciento. Posteriormente, pensando en sistemas de órdenes superiores, algunas administraciones ensayaron la utilización de cables coaxiales de muy pequeño diámetro (microaxiales), que han dado buen resultado para sistemas de hasta tercer orden. Desde luego, la utilización de coaxiales convencionales de pequeño diámetro y diámetro normal, empleados tradicionalmente para sistemas analógicos, es posible para sistemas MIC, y en ellos se están ensayando sistemas de hasta quinto orden. Por último, las fibras ópticas, portador de consideración reciente, nunca utilizado en técnica analógica, parece ofrecer ilimitadas posibilidades para la técnica digital.

Hasta aquí se ha pretendido dar una idea sobre las peculiaridades de la técnica MIC, insistiendo en la estructura conceptual de la misma. Para completar el panorama, conviene pasar revista a las principales características operativas de di-

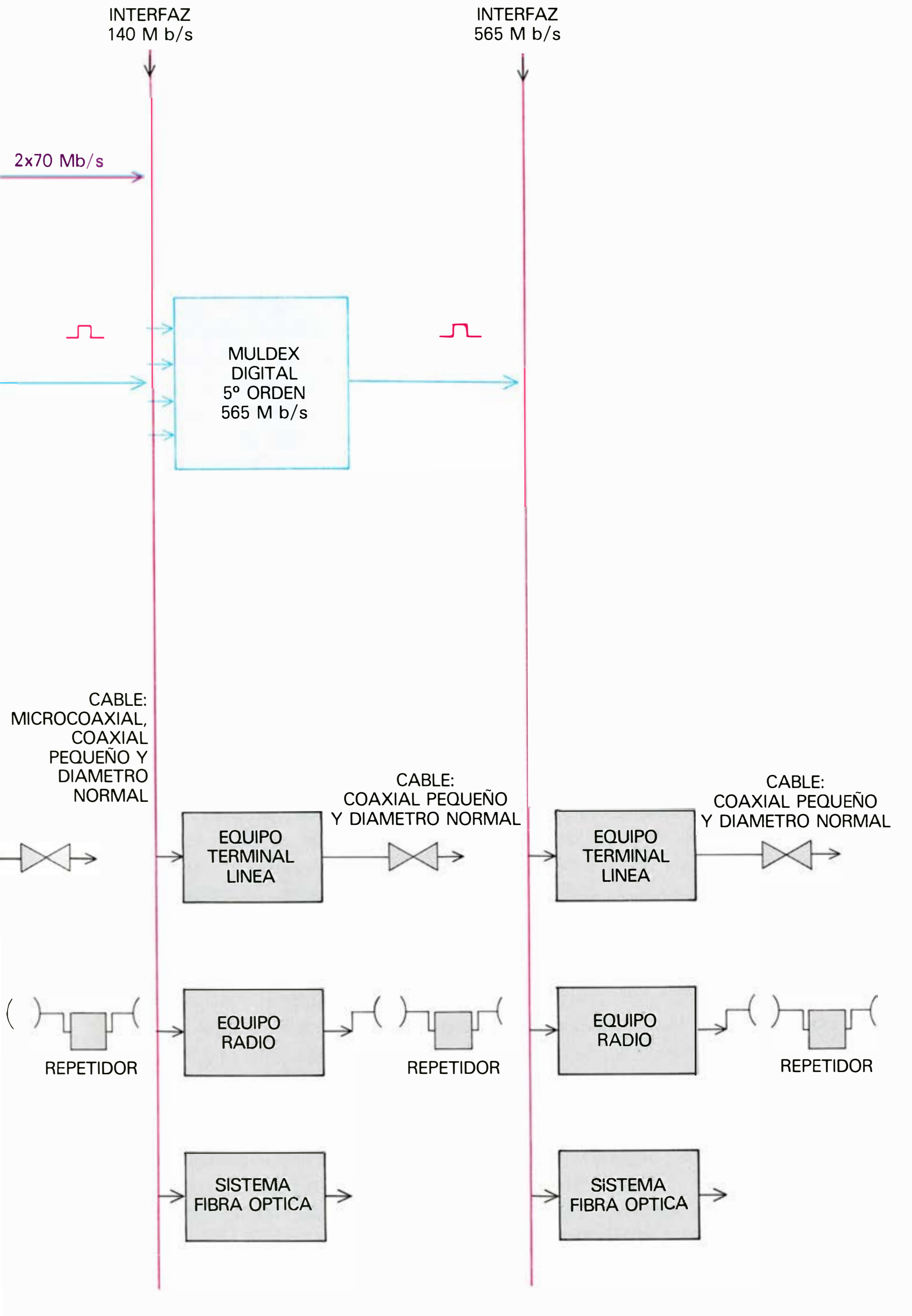


JERARQUIA MIC EUROPEA. En el gráfico se representan los diferentes equipos de la jerarquía MIC, incluyendo otros servicios distintos de la telefonía.

En el artículo se han expuesto las técnicas digitales de transmisión basadas en la modulación por impulsos codificados (MIC), partiendo de su justificación a la

cha técnica, características que, bajo la forma de ventajas o inconvenientes de la técnica digital respecto de las analógicas, define el grado de interés y de expansión de aquélla. Las ventajas o inconvenientes de una técnica, bajo los aspectos operativos, hay que verlos necesariamente a la luz de consideraciones técnicas y económicas. Es decir, una técnica será competitiva y, por tanto, de aplicación extendida, si proporciona las mismas prestaciones que otra con menor coste, o si ofrece mayores posibilidades a un coste similar.

Hay tres características que marcan las ventajas fundamentales de las técnicas digitales respecto a las analógicas. La primera es la relativa inmunidad respecto al ruido e interferencias de una transmisión digital. En efecto, la naturaleza de las señales digitales hace que baste con el reconocimiento de la existencia de un impulso, para regenerarlo, no importando, dentro de amplios márgenes, la degradación que ese impulso haya podido sufrir durante la transmisión. Ello conduce a que la calidad resulte prácticamente independiente de la distancia recorrida. Puesto que el ruido se debe casi de un modo exclusivo al proceso de cuantificación, la calidad depende sólo del equipo terminal, no de la línea, y el terminal se puede diseñar con la calidad que se precise. En la técnica analógica, en cambio, es preciso cuidar las señales transmitidas en cada punto de su recorrido, ya que el ruido y las interferencias que inciden en ellas durante ese recorrido quedan asociadas a las magnitudes que llevan la información, no pudiéndose separar de ella luego, y acumulándose su efecto.



luz de la teoría de la información y presentando su desarrollo y su incidencia en las telecomunicaciones. ***Muldex*** **es la denominación que designa, por contracción, los procesos de multiplexación y demultiplexación.**

La segunda característica importante se refiere a la utilización de la banda de frecuencias. La transmisión de una señal compuesta por una sucesión de impulsos exige una anchura de banda bastante mayor que la que necesita una señal analógica que contenga la información del mismo número de canales telefónicos. Claro está, esto no es una ventaja de la técnica digital, sino todo lo contrario, al menos aparentemente. Sin embargo, sucede, en línea con lo que se acaba de mencionar en los párrafos precedentes, que la calidad exigida por la técnica analógica a lo largo de toda la banda utilizada es muy alta, y hay que cuidar dicha banda con esmero en todas y cada una de sus partes. Por el contrario, la transmisión digital no requiere tales cuidados; es muy poco exigente en razón de que cualquier deformación, aunque sea importante, de los impulsos no produce efecto significativo. Y como consecuencia, resulta mucho más fácil (y, por tanto, económico) proveer la mayor anchura de banda precisa para una señal digital sin requisitos especiales de calidad, que la anchura de banda más reducida, pero en condiciones mucho más estrictas, necesaria para una señal analógica equivalente.

Que esto constituya una importante ventaja lo demuestran algunos ejemplos simples: una señal de videotelefonía puede transmitirse mediante una señal digital de 8448 kilobits por segundo (segundo orden), que equivale a 120 canales telefónicos en la norma europea MIC. Esa misma señal, en forma analógica, necesita de una banda de frecuencias que, aún siendo inferior, es la equivalente a 300 canales telefónicos analógicos. De la misma forma, para un programa de televisión en color, puede pensarse en una señal digital de 34.368 kilobits por segundo (terciaria) equivalente a 480 canales telefónicos, mientras que, en forma analógica, equivale a 960 canales.

Por último, la tercera característica estriba en que la técnica digital permite unificar la naturaleza y comportamiento de la red para cualquier tipo de información (voz, vídeo o datos), obteniéndose una mayor eficacia global, en particular si se tie-

nen en cuenta los ejemplos mencionados en el párrafo precedente. Estas tres características representan ventajas importantes. Ahora bien, tales ventajas técnicas no serían significativas si no fuesen acompañadas de consideraciones económicas también positivas. Las tecnologías actuales de fabricación permiten a los sistemas digitales obtener esas ventajas económicas.

Los circuitos digitales son más complejos que los analógicos. Sin embargo, las tecnologías de fabricación de circuitos integrados permiten obtener grandes series a precios reducidos. Y sucede que las tolerancias de fabricación de esos circuitos para técnica analógica son relativamente estrechas, ya que cualquier variación en su comportamiento tiene una repercusión significativa sobre las señales. Por el contrario, en los circuitos digitales cuyo funcionamiento es de "todo o nada", las variaciones en su comportamiento, dentro de márgenes amplios, no tienen incidencias sustanciales y sus tolerancias son mayores. Esto abarata de forma importante la fabricación.

En resumen, las principales ventajas de los sistemas digitales pueden centrarse en los siguientes puntos: (*1*) Ofrecen una calidad independiente de la distancia, homogénea de unos canales a otros, y de unas rutas a otras; lo cual facilita, por otra parte, la planificación de redes. (*2*) Permiten un tratamiento homogéneo de las señales con independencia de la naturaleza de la fuente, lo que aumenta la eficacia de la red y facilita, asimismo, la planificación, ya que pueden integrarse en una misma red diversas prestaciones o servicios. En esta idea de integración incide de forma sustancial y positiva la tendencia a utilizar sistemas de conmutación temporales, que se basan en técnicas digitales, lo que aumentará la homogeneidad de la red y, por tanto, su eficacia y economía. (*3*) El costo de los sistemas digitales no es intrínsecamente superior al de los analógicos, e incluso puede llegar a ser inferior al de éstos.

Como contrapartida de estas ventajas, cabe mencionar los siguientes inconvenientes principales: (*1*) Las redes actuales son, en su mayoría (y hasta totalmente en algunos casos) de naturaleza analógica. Esto representa una desventaja inicial a la hora de empezar a instalar sistemas digitales. La compatibilidad entre ellos y los analógicos hay que resolverla mediante interfaces o adaptaciones que recargan el coste de los digitales, que son los que tienen que acoplarse a lo que ya está. (*2*) Por su reciente expansión, aún están poco desarrollados y extendidos los sistemas digitales de órdenes superiores. En conjunto, la fabricación de sistemas digitales no alcanza los volúmenes que todavía tienen los analógicos. Ello provoca que sus costes no estén tan optimizados, y, en consecuencia, que la competencia económica sea crítica en muchos casos. (*3*) Debido a los procesos de codificación-decodificación de señales, existe una limitación en cuanto al número de codecs (codificadores-decodificadores) que pueden conectarse en serie en una comunicación. De estos tres inconvenientes, los dos primeros son circunstanciales y deben desaparecer con el tiempo. En cuanto al tercero, también es transitorio, aunque quizá no se aprecie a primera vista. Sin embargo, es fácil ver que la necesidad de incluir más de una pareja de codec en una comunicación aparece cuando se da la coexistencia de áreas de la red con tecnología analógica y con tecnología digital. Cada vez que, a lo largo de una comunicación, se pasa de un medio de transmisión analógico a otro digital o viceversa, es preciso insertar un codec. Si la red fuese totalmente digital, sólo sería necesario un codec al principio y otro al final de una comunicación. Por tanto, a medida que la digitalización se vaya extendiendo, la necesidad de insertar más de una pareja de codec disminuirá. Esta faceta tiene también carácter circunstancial. En cualquier caso, no se trata de una limitación grave, y con una adecuada planificación puede resolverse sin dificultad aun en el peor de los casos.

De todo lo expuesto se desprende que la futura expansión de los sistemas digitales está asegurada, y de hecho la realidad actual abona esta afirmación, ya que la tendencia en todas las administraciones telefónicas apunta indefectiblemente hacia una progresiva "digitalización" de la red. La situación mundial en estos momentos, en cuanto a utilización de sistemas digitales MIC, es como sigue. Los sistemas de primer orden abrieron la marcha, y son, en consecuencia, los más extendidos en todo el mundo. En Europa, los países que cuentan con mayor número de sistemas primarios MIC instalados son Italia, Francia y España. En nuestro país existen actualmente en explotación unos mil sistemas de 2048 Kb/s. En el marco de los países miembros de la CEPT, se ha potenciado el desarrollo de cables especiales para transmisión de sistemas MIC, tanto de pares como coaxiales de muy pequeño diámetro (microcoaxiales). Los primeros permiten la transmisión de sistemas primarios principalmente, y los segundos la de sistemas secundarios y terciarios. Hasta el mencionado nivel terciario, existen sistemas operativos en varios países europeos. Los de órdenes superiores a ése se encuentran a nivel de experimentación. Sobre coaxiales convencionales de pequeño diámetro se están experimentando las velocidades de tercero, cuarto y quinto orden, y para estos dos últimos órdenes se ensaya, asimismo, el coaxial de diámetro normal. En el campo de los radioenlaces digitales se piensa en los mismos órdenes jerárquicos mencionados y algunos no jerárquicos, en las bandas de 10 a 19 gigahertzs, bandas nunca utilizadas para radioenlaces analógicos. En Norteamérica (Estados Unidos y Canadá) y en Japón, existen sistemas comerciales y en servicio hasta el cuarto orden inclusive, hallándose en fase de experimentación los de quinto.

Por último, cabe mencionar las fibras ópticas, como portador inédito hasta la fecha, y que parecen ofrecer interesantes perspectivas cara a un futuro a medio o largo plazo. En ellas se ensayan todos los niveles jerárquicos, aunque se prevé que su más eficaz aplicación estará en los órdenes superiores al cuarto.

El advenimiento de las técnicas de transmisión digitales, y especialmente la técnica MIC, ha supuesto una auténtica revolución en el mundo de las telecomunicaciones. Desde la invención de la válvula termoiónica, hasta la fecha, toda la evolución de las redes de telecomunicación ha tenido lugar dentro de los cauces analógicos. Los sucesivos avances tecnológicos, que representaron escalones importantes en el perfeccionamiento y desarrollo de los equipos de transmisión (el paso del tubo de vacío al transistor, y de éste a los circuitos integrados), han permitido alcanzar cotas de refinamiento técnico increíbles hace algunos lustros.

Sin embargo, la adopción de las técnicas digitales, más que un nuevo escalón, ha supuesto ahora el inicio de una nueva era, por cuanto se modifica la naturaleza misma de las señales eléctricas que se manejan, y como consecuencia, obligará a revisar todos los criterios de planificación y concepción de las redes de telecomunicación. La digitalización de las redes se ha iniciado ya, y su progresión no ofrece dudas. A ello contribuirá, de forma definitiva, la adopción, también iniciada, de centrales de conmutación temporales que utilizan técnicas digitales. Evidentemente, las técnicas analógicas convencionales seguirán presentes aún mucho tiempo, coexistiendo con las digitales, dada la gran inercia que las redes de telecomunicación presentan a causa de las fabulosas inversiones que suponen.

Pero en un futuro más o menos lejano, de incierta cronología, pero de segura predicción, las redes de telecomunicación alcanzarán una naturaleza plenamente digital, y su estructura conceptual, como consecuencia de ello, estará muy alejada de la que han venido teniendo hasta la fecha.

Función cerebral y flujo sanguíneo

Los cambios en la cantidad de sangre que fluye por diferentes áreas de la corteza cerebral, y que denotan variaciones en la actividad de las mismas, se revelan de una forma gráfica con la ayuda de un isótopo radiactivo

Niels A. Lassen, David H. Ingvar y Erik Skinhøj

Los frenólogos de principios del siglo XIX creían que las cualidades morales de un individuo, como la "creatividad" y la "agresividad", podían deducirse a partir del molde de las protuberancias craneales, que reflejarían presumiblemente una hipertrofia de regiones específicas del cerebro. Aunque ese sagaz y persistente punto de vista está en absoluto descrédito en la actualidad, contenía, paradójicamente, un germen de verdad. Los estudios de los efectos de las lesiones cerebrales en el comportamiento y de las estimulaciones eléctricas del cerebro durante intervenciones quirúrgicas han revelado que las funciones sensitivas y motoras específicas (más que las características de la personalidad) se localizan en regiones delimitadas de la corteza cerebral, capa de cinco milímetros de células nerviosas que constituye la superficie del cerebro. En nuestro laboratorio del Hospital Bispebjerg de Copenhague y en la Universidad de Lund en Suecia, una nueva técnica de radioisótopos nos ha permitido observar directamente la localización de la función de la corteza cerebral humana.

Los primeros intentos serios por localizar las funciones en la corteza fueron llevados a cabo, en 1861, por el médico francés Paul Broca, quien realizó autopsias a dos paralíticos; ambos habían tenido afectado únicamente el lado derecho y habían sufrido una afasia grave con pérdida del habla. En las autopsias Broca encontró graves lesiones del tejido cerebral en parte del hemisferio izquierdo y dedujo que la función del lenguaje se localiza en esa región del cerebro. Posteriormente, los investigadores utilizaron la técnica indirecta de Broca para localizar muchas otras funciones corticales basándose en los trastornos neurológicos y mentales resultantes de lesiones del cerebro.

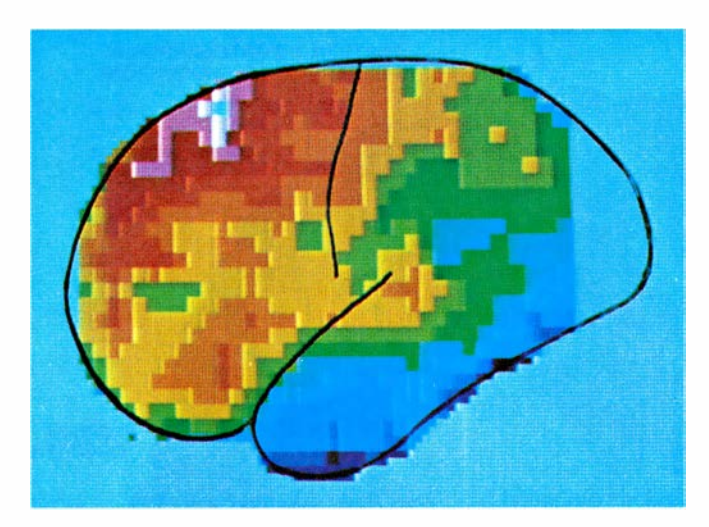
En las décadas de 1940-50 y 1950-60 varios neurocirujanos, en particular Wilder Penfield y sus colaboradores, del Instituto Neurológico de Montreal, topografiaron las áreas corticales mediante la estimulación eléctrica de la corteza durante intervenciones quirúrgicas que requerían, por la naturaleza de la enfermedad subyacente, la exposición de la superficie del cerebro. Las características de la respuesta a cada estimulación se determinaban a partir del informe verbal del paciente, que permanecía consciente durante toda la intervención. Por ejemplo, al estimular puntos del área somatosensorial, el paciente daba cuenta de sensaciones de entumecimiento o de cosquilleo en regiones específicas de la superficie corporal.

Estos estudios de localización, apoyados en numerosas observaciones en animales, dieron lugar al establecimiento de un campo de relaciones básicas entre las áreas corticales y la función: el lóbulo occipital, que se halla en la parte posterior de la corteza, interviene en la visión; la corteza postcerebral desempeña un papel en la sensibilidad dérmica y muscular; partes de los lóbulos temporales participan en la audición; la corteza precentral contribuye al control motor y las áreas de Wenicke y de Broca (que en casi todas las personas se localizan en el hemisferio cerebral izquierdo) están implicadas, respectivamente, en la comprensión y en la síntesis del lenguaje. Existen otras áreas relativamente amplias de la corteza que ni son activadas por los estímulos sensoriales ni contribuyen al desencadenamiento de la actividad motora; estas regiones, que integran las llamadas áreas de asociación, intervienen en el procesamiento detallado de los estímulos sensoriales, en la planificación de las acciones motoras y quizás en funciones puramente mentales.

Aunque los estudios basados en las lesiones cerebrales y la estimulación eléctrica han suministrado mucha información sobre la localización de las funciones en la corteza, estas experiencias tienen una limitación inherente. Como fue constatado ya en 1874 por John Hughlings Jackson, del National Hospital for Nervous Diseases de Londres, "no es lo mismo localizar el lenguaje que la lesión que destruye el lenguaje". Por ejemplo, como se verá en este artículo, el hemisferio derecho es mucho más activo durante el habla de lo que sugiere el estudio basado en las lesiones de los tejidos. Un enfoque más directo del problema de la localización es el que se basa en el estudio de la corteza cerebral intacta en diferentes situaciones funcionales, a saber: mientras "el cerebro trabaja". Una técnica consiste en registrar con microelectrodos las alteraciones de las descargas de impulsos en células discretas de la corteza. Sin embargo, incluso muchas dotaciones de microelectrodos resultan insuficientes para descifrar el amplio complejo desplegado por los 10.000 millones de células nerviosas que hay en el cerebro. Hemos adoptado un enfoque completamente distinto y más global: la medición de la intensificación del riego sanguíneo de las áreas corticales que se produce durante el desempeño de actividades sensoriales, motoras y mentales específicas.

Nuestro método se basa en el hecho de que el flujo a través de los tejidos del cuerpo varía con la intensidad del metabolismo y el grado de actividad funcional de los mismos. La interrelación entre metabolismo enegético y flujo sanguíneo fue postulada por primera vez en 1914 por Joseph Barcroft, de la Univeridad de Cambridge; señaló que sólo puede mantenerse un nivel funcional más elevado de un tejido a expensas de un aumento en la intensidad de consumo de oxígeno. La mayor parte de la energía utilizable por el cuerpo procede del desdoblamiento de la molécula rica en energía de adenosín trifosfato (ATP) en los productos de baja energía: adenosín difosfato (ADP) y fosfato inorgánico (P_i). El ATP se vuelve a constituir luego a partir de los productos procedentes de su fragmentación mediante una reacción que requiere oxígeno y glucosa (fosforilación oxidativa).

Como se mantiene una proporción constante entre el número de moléculas de ATP regeneradas y el número de moléculas de oxígeno utilizadas en el proceso, la intensidad funcional de un tejido está es-



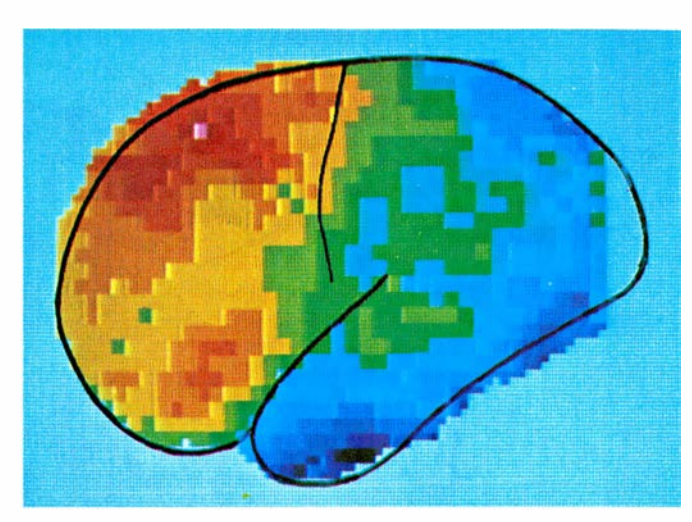
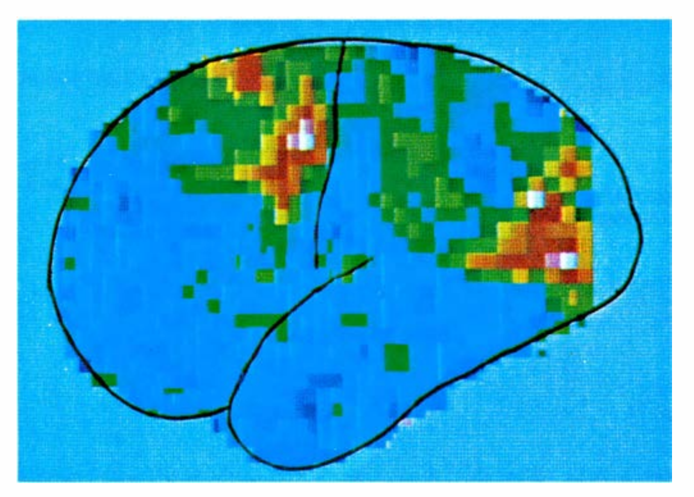
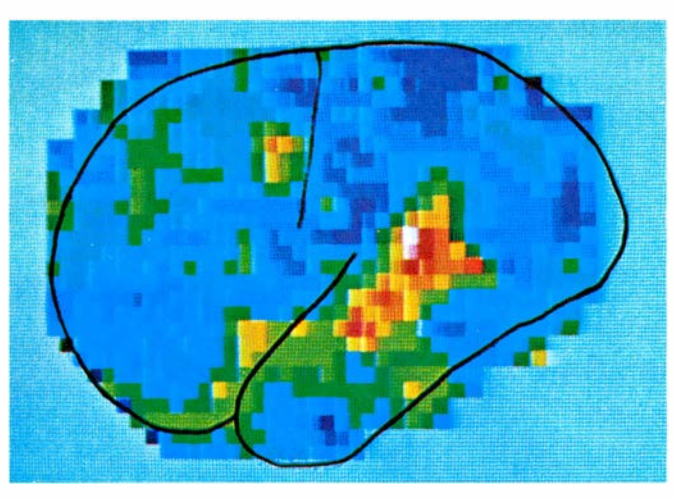
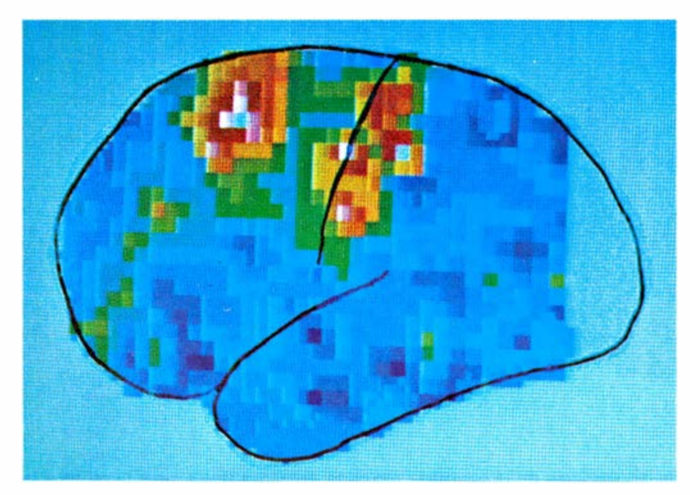


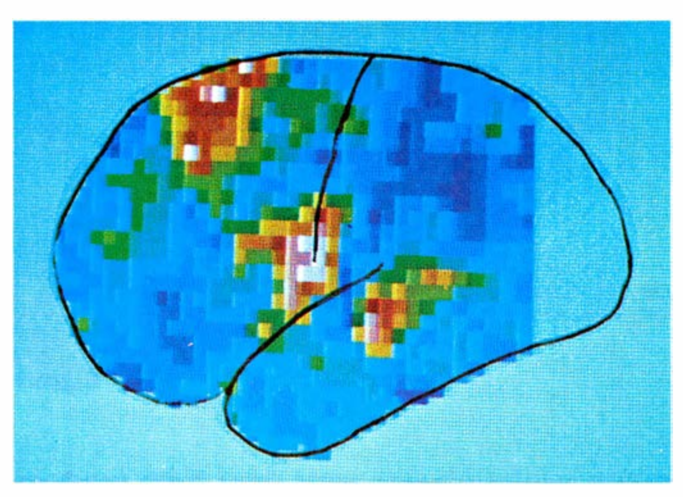
IMAGEN NORMAL EN REPOSO de la actividad celular nerviosa en los hemisferios izquierdo y derecho obtenida midiendo el flujo sanguíneo regional, que está estrechamente relacionado con la tasa metabólica y, por tanto, con la actividad funcional. Las imágenes se formaron en un ordenador a partir de los datos obtenidos mediante la detección del paso del isótopo radiactivo xenón 133 a través de la corteza. Cada pixel, o elemento de la figura, representa un centímetro cuadrado de corteza. En la escala de colores la tasa media de flujo es verde, tasas hasta el 20 por ciento por debajo de la media están sombreadas en azul y tasas hasta el 20 por ciento por encima de la media en rojo. Las imágenes sugieren que en estado de reposo las áreas frontales presentan gran actividad.



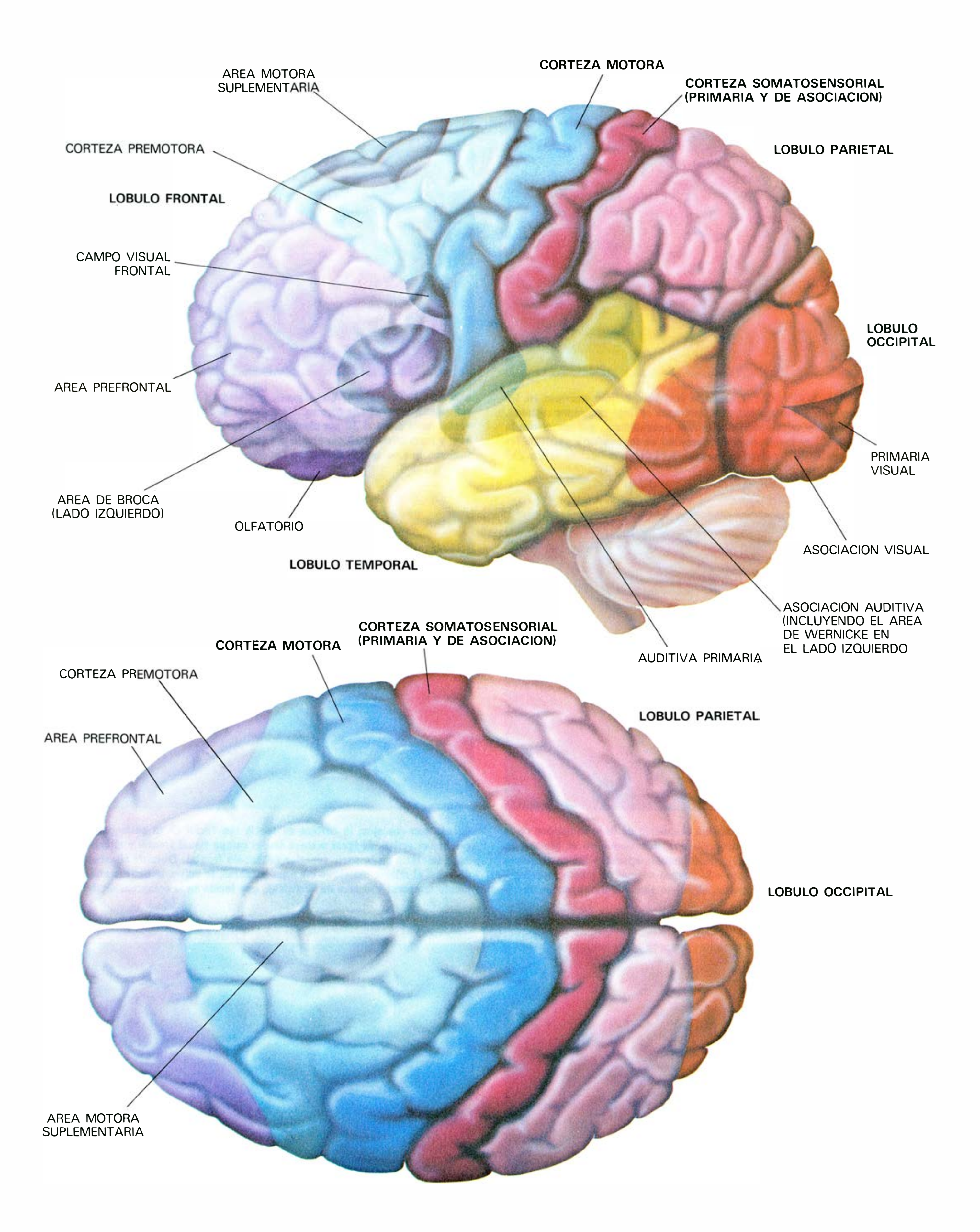


LA PERCEPCION SENSORIAL cambia el modelo de flujo sanguíneo en la corteza, revelando la localización de áreas que intervienen en el proceso de información sensorial. En estas imágenes solamente aparecen las desviaciones de la intensidad de flujo con respecto a la imagen normal en reposo hiperfrontal. Durante el estudio de la izquierda, el sujeto siguió a un objeto en movimiento con sus ojos; la corteza de asociación visual en la parte posterior del cerebro se activa, de igual manera que el campo visual frontal y el área motora suplementaria en la parte superior del lóbulo frontal. Durante el estudio de la derecha, el mismo individuo oía unas palabras; se activa la corteza auditiva del lóbulo temporal y el área de Wernicke, que incide en la comprensión del habla.





MOVIMIENTO VOLUNTARIO, que activa diferentes partes de la corteza central de acuerdo con la parte del cuerpo implicada. Durante el estudio de la izquierda, el sujeto movió los dedos del lado del cuerpo opuesto al hemisferio examinado, activándose el área de los dedos de la mano en la corteza central y el área motora suplementaria. Durante el estudio de la derecha, se le dijo al individuo que moviera su boca contando hasta 20 de forma continua, activándose el área bucal de la corteza central, el área motora suplementaria y la corteza auditiva. Niels A. Lassen sacó todas estas imágenes por ordenador.

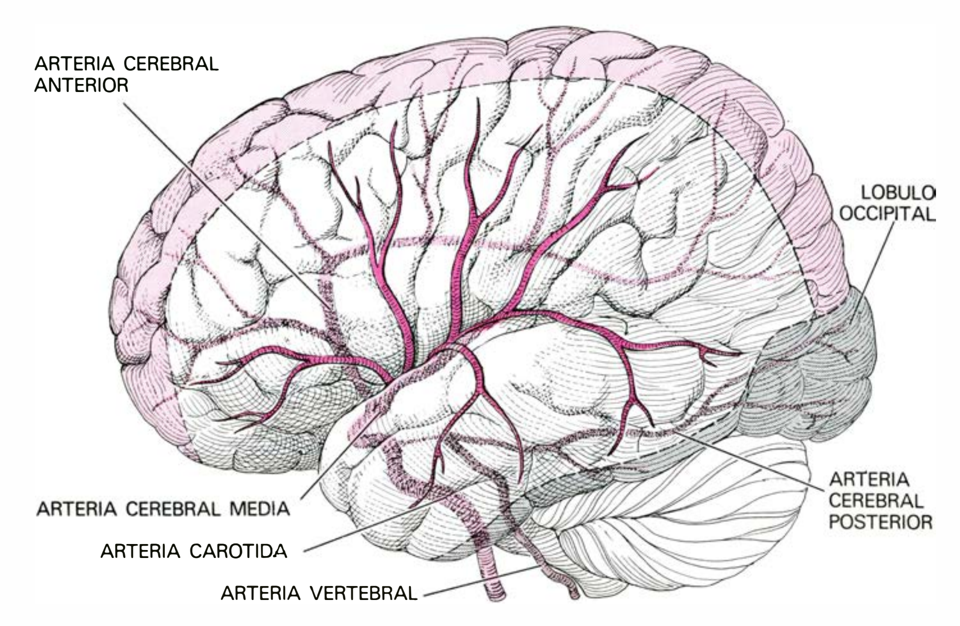


MAPA FUNCIONAL de la corteza cerebral humana obtenido mediante una combinación de métodos: estudios de los efectos del deterioro del cerebro, estimulación de la corteza expuesta durante la cirugía cerebral, registros mediante microelectrodos de la actividad cortical motivada por estímulos de comportamiento o sensoriales y por la técnica de isótopos radiactivos descrita en este artículo. Cada modalidad sensorial tiene un área cortical primaria que recibe las proyecciones de la superficie receptora periférica (tales como la retina o la piel) y un área de asociación que procesa la llegada sensorial y la almacena en una serie de memorias específicas. El área premotora está implicada en actividades motoras complejas tales como la de escribir a máquina. El area motora suplementaria programa los movimientos voluntarios secuenciales. El área de Broca controla la coordinación de los músculos relacionados con el habla.

trechamente asociada a su consumo de oxígeno. El oxígeno se suministra a los tejidos por medio de la circulación, y un aumento en la demanda de oxígeno es cubierto por un flujo más elevado de sangre oxigenada. Durante la actividad muscular intensa, el flujo a través del tejido muscular puede incrementarse hasta 20 veces sobre el nivel normal; se ha comprobado que la intensidad circulatoria a través de otros órganos puede ser también ajustada minuciosamente según las variaciones en la demanda de oxígeno. Como esta precisión en el ajuste del flujo se mantiene en órganos aislados y completamente denervados, parece que debe existir un control mediante los factores químicos liberados en los procesos metabólicos locales, si bien no se comprende todavía el mecanismo preciso de este proceso.

En el caso del cerebro, la interrelación entre la actividad funcional, metabolismo y flujo sanguíneo fue sugerida por primera vez en 1890 por Charles S. Roy y Charles S. Sherrington, quienes, trabajando en el Laboratorio de Patología de Cambridge, observaron que segundos después del inicio de un ataque de epilepsia (hiperactividad sincrónica de células cerebrales manifestada a través de convulsiones musculares) se producía una turgencia del cerebro; ello sugería que se daba una elevación masiva en el suministro de sangre.

Investigaciones subsiguientes han revelado que el metabolismo cerebral y el flujo sanguíneo aumentan también durante la activación normal de las regiones de la corteza cerebral que intervienen en el desarrollo de las funciones específicas. En 1937, Carl F. Schmidt y James P. Hendrix, de la Facultad de Medicina de la Universidad de Pennsylvania, registraron mediante termopares un incremento, estrictamente localizado, del flujo sanguíneo a través de la corteza visual de un gato, cuando dirigieron un haz puntual de luz hacia la retina del animal. Desde entonces se han descrito otros muchos ejemplos de incrementos localizados en el flujo sanguíneo cerebral en animales y en el hombre; hoy se ha puesto ya de manifiesto que durante un test que implique el desarrollo de un tipo específico de función por el cerebro existe un cambio local en la actividad de las células nerviosas y, por consiguiente, en la actividad metabólica, que a su vez da lugar a un incremento del flujo sanguíneo en la región activa. Este descubrimiento ha permitido la localización de funciones cerebrales a través del estudio de las variaciones regionales del flujo circulatorio. Puede obtenerse la misma información estudiando la distribución de los cambios metabólicos locales (midiendo el consumo de oxígeno o de glucosa), pero hasta el momento la localización funcional en la corteza humana sólo se ha estudiado de forma sistemática a través de mediciones de flujo sanguíneo.



PRINCIPALES ARTERIAS que irrigan la corteza cerebral, esquematizadas. Las regiones del habla y de la audición son abastecidas por la arteria cerebral media. Las regiones frontal y del surco anteroposterior están irrigadas por la arteria cerebral anterior. Ambas arterias nacen de la arteria carótida en el cuello. El área visual primaria en la parte posterior de la corteza, sin embargo, está irrigada por la arteria cerebral posterior, la cual procede de la arteria vertebral en el cuello. Debido a que el xenón radiactivo se inyecta usualmente en la carótida, la corteza visual primaria no aparece en las imágenes de flujo sanguíneo. La interrelación entre actividad cerebral y flujo sanguíneo fue sugerida por Charles S. Roy y Sherringhton, de Cambridge.

El flujo cerebral medio fue determinado por primera vez en 1944 por Seymour S. Kety, quien a la sazón trabajaba con Schmidt. El método de Kety consistió en observar la velocidad a que el cerebro se saturaba o desaturaba con un gas inerte. El sujeto inhalaba una mezcla de gases, que contenía un 15 por ciento de óxido nitroso, durante 10 minutos, a lo largo de los cuales se controlaba la concentración del gas en muestras de sangre arterial y venosa procedente del cerebro. El área comprendida entre las curvas de saturación arterial y venosa proporcionaba una medida del flujo sanguíneo medio, cuyo valor normal es aproximadamente de 50 mililitros de sangre por cada 100 gramos de tejido cerebral y por minuto. Kety determinó también la diferencia del contenido de oxígeno entre la sangre arterial y venosa, que es normalmente de unos 0,7 mililitros de oxígeno por mililitro de sangre. Multiplicando el flujo por la diferencia de oxígeno se obtiene la intensidad media de consumo metabólico de oxígeno cuyo valor normal es de unos 3,5 mililitros de oxígeno por 100 gramos de tejido y por minuto.

Posteriormente se pudo medir el flujo sanguíneo en regiones circunscritas del cerebro humano intacto con la ayuda de isótopos radiactivos. Esta técnica la desarrollamos nosotros tres, en 1961, en estudios llevados a cabo con gatos; su aplicación se extendió muy pronto a pruebas de diagnóstico en el hombre. El método implica la utilización de xenón 133, un isótopo radiactivo del gas inerte xenón. Se disuelve el gas radiactivo en suero fisiológico estéril y se inyecta rápidamente un pequeño volumen (de dos a tres mililitros que contienen 3-5 milicuries de radiactividad) en una de las principales arterias que irrigan el cerebro. La llegada y desaparición subsiguiente de la radiactividad en muchas regiones del cerebro se controla durante un minuto mediante una cámara de rayos gamma integrada por una dotación de 254 detectores de centelleo situados externamente, cada uno de los cuales se ajusta mediante el colimador para explorar un centímetro cuadrado de superficie cerebral, más o menos. La información procedente de los detectores se procesa con un pequeño ordenador digital y se proyecta en un monitor de televisión en color; a cada intensidad de flujo corresponde un color o matiz diferente. Debido a la atenuación de la radiación procedente de las estructuras más profundas del cerebro, la radiación gamma detectada procede de la corteza superficial cerebral. En consecuencia, la técnica del xenón radiactivo proporciona una imagen auténticamente específica de la actividad de la corteza cerebral que subyace a los puntos de distribución de los electrodos.

Venimos utilizando esta técnica de forma rutinaria en pacientes a los que se les debe practicar una arteriografía cerebral. La arteriografía cerebral comporta la realización de una radiografía de los vasos sanguíneos cerebrales después de la inyección de un medio opaco a los rayos X en una arteria. Esta inyección se realiza a través de un catéter (tubo fino de plástico)

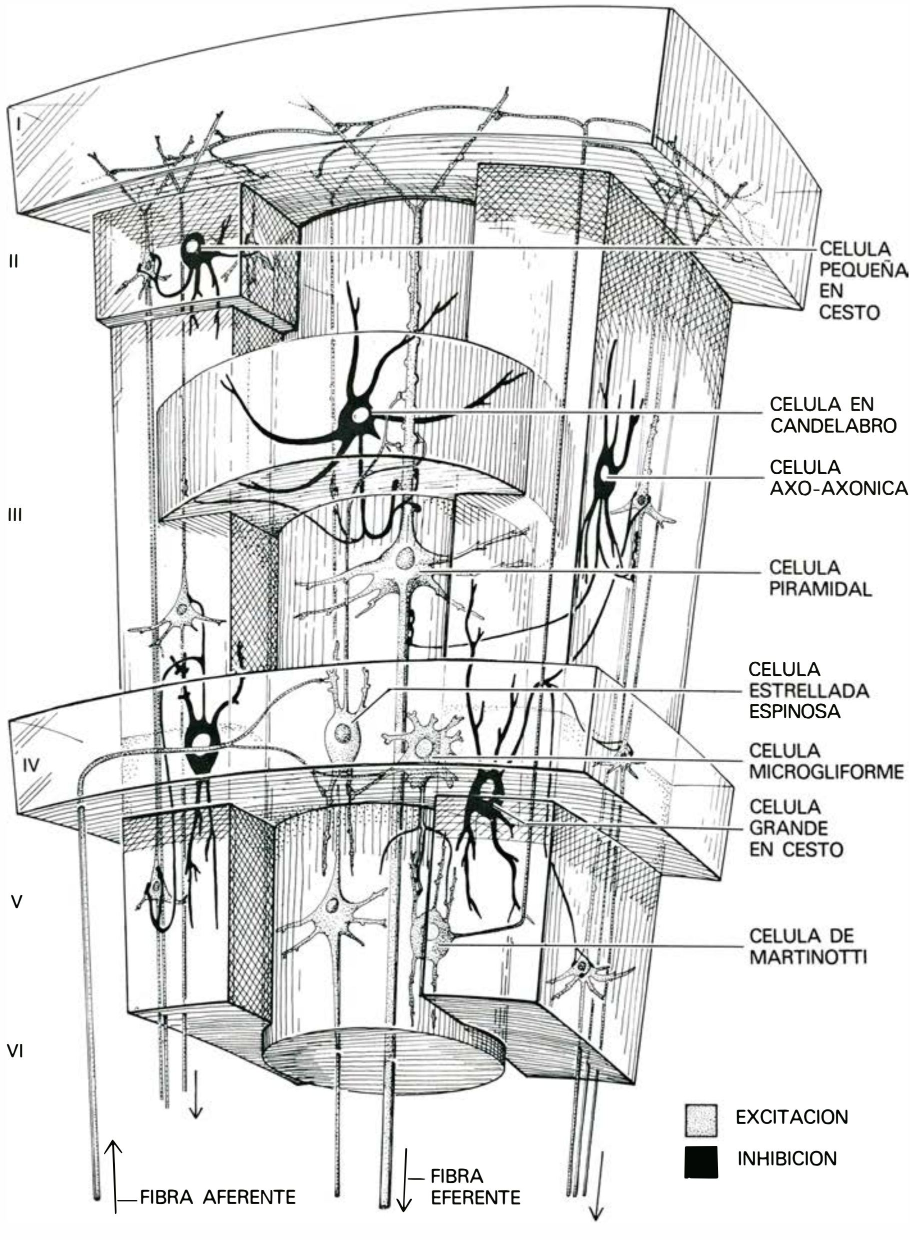
colocado directamente en la arteria carótida interna mediante punción en el cuello o alcanzándola a través de la arteria femoral por medio de punción en la ingle. Como el mismo catéter se emplea para la inyección del xenón radiactivo, las medidas del flujo cerebral no acarrean un riesgo especial para el paciente. (La baja intensidad de la radiación gamma emitida por el xenón 133 no se considera peligrosa.)

En nuestros laboratorios de Copenhague y Lund sometimos a unos 500 pacientes a examen con la técnica del xenón radiactivo con fines diagnósticos en casos de apoplejía, tumores o epilepsia. En visión retrospectiva, podemos afirmar que el 80 por ciento de estos pacientes tenían un cerebro normal en el momento de su examen. Este grupo estaba constituido por pacientes aquejados de fuertes migrañas, ataques generalizados de epilepsia y otros síntomas neurológicos transitorios que cesaron y no fueron atribuidos a lesiones o anormalidades cerebrales permanentes. La electroencefalografía y otros tests sirvieron también para confirmar la normalidad de sus cerebros.

Nuestros estudios sobre el flujo cerebral local en estos pacientes nos permitieron obtener así algunas conclusiones sobre la localización de las funciones en la corteza cerebral normal

La pauta habitual del flujo sanguíneo en el cerebro de un individuo normal en reposo es muy característica y reproducible; además sirve de punto de partida para la interpretación de las imágenes funcionales registradas durante diferentes tipos de actividad sensorial, motora o sólo mental. Sorprendentemente, cuando un individuo está despierto y en reposo, en cómoda posición supina con los ojos cerrados y el laboratorio en silencio, sin que nadie hable ni le toque, la imagen que ofrece el flujo sanguíneo a través de la corteza no es uniforme. Al contrario, el flujo permanece siempre substancialmente más elevado en la región frontal de la corteza que en la central y posterior. Aunque hemos mencionado que el flujo medio es de 50 mililitros por 100 gramos de tejido cerebral y por minuto, el de la parte frontal es de un 20 a un 30 por ciento superior a este promedio; en correspondencia, el flujo de las partes posteriores, en particular el de algunas regiones del lóbulo temporal, es inferior. La densidad tanto de capilares como de células nerviosas en las distintas regiones de la corteza es semejante; consecuentemente, la notable diferencia que se advierte en la intensidad de flujo sugiere que el grado de actividad general de la parte frontal del cerebro en condiciones de reposo es un 50 por ciento superior al grado de actividad de las partes posteriores.

Esta pauta de un elevado flujo frontal en condiciones de reposo puede contribuir a la comprensión de la autoconsciencia. Es sabido que las partes más anteriores del lóbulo frontal, las áreas prefrontales, son responsables de la planificación de la conducta en su más amplio sentido, mientras que las restantes regiones de la corteza cerebral ejercen funciones sensoriales y motoras. La existencia de un flujo frontal intenso sugiere en consecuencia que en estado de vigilia y reposo el cerebro trabaja en la planificación y selección de diferentes normas de comportamiento. En el mismo estado, las regiones motora y sensitiva de la corteza no denotan gran actividad, quizás incluso sean objeto de inhibición. Esta interpretación parece estar de acuerdo con nuestra experiencia subjetiva. Mientras uno permanece en reposo no es consciente de forma continua de todos los estímulos sensoriales: sólo ocasionalmente percibe distintas señales visuales, auditivas o táctiles que destacan del "ruido" de fondo propio del estado de reposo. La mayor parte de la actividad consciente en condiciones de reposo se dedica a pensamientos internos, particularmente a reflexiones sobre la propia situación actual y sus relaciones con los acontecimientos pasados y los posibles futuros. Puede, pues,



ARQUITECTURA CELULAR de la corteza cerebral organizada en módulos columnares con los circuitos de las células nerviosas dispuestos artificialmente en vertical. Una columna típica, de unos 250 micrometros (10^{-3} milímetros) de diámetro se muestra en este diagrama considerablemente simplificado, basado en otro de János Szentágothai, de la Universidad Semmelweis de Budapest. Cientos de fibras nerviosas aferentes que transportan información sensorial convergen en las células estrelladas espinosas en la Capa IV de la corteza. Los circuitos verticales de interneuronas, dispuestos según una configuración espacial muy específica, transforman estos datos inconexos en la sutilezas de la experiencia consciente y del comportamiento humano.



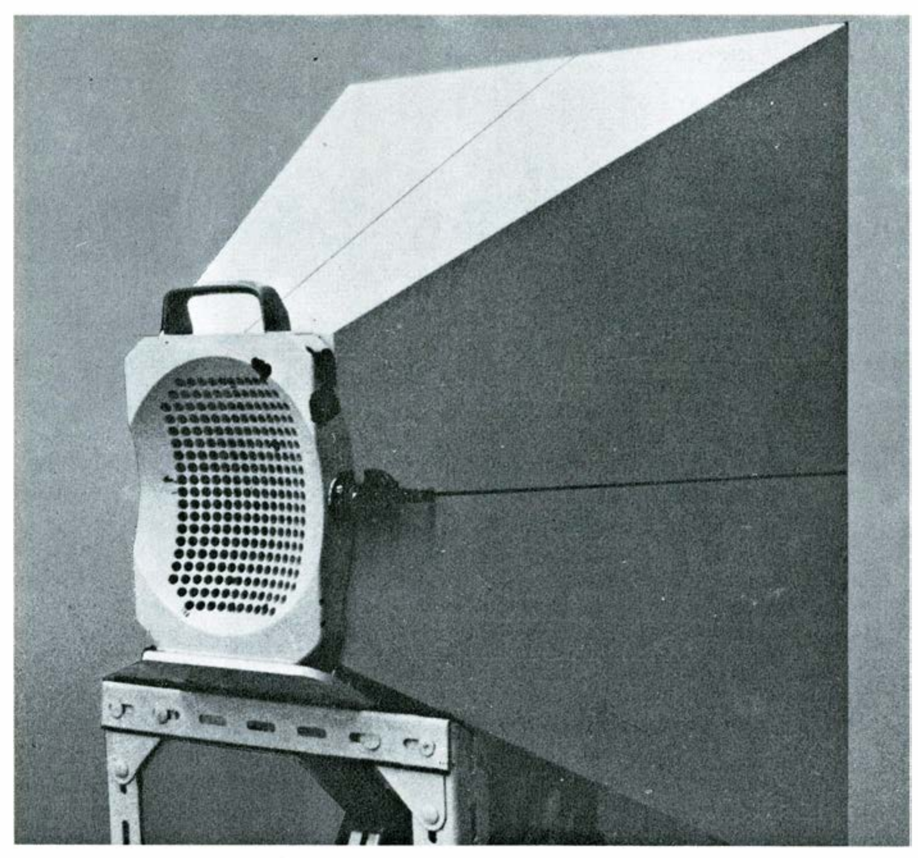
APARATO DE EXPLORACION CEREBRAL empleado por los autores en sus laboratorios en el Hospital Bispebjerg de Copenhague y en la Universidad de Lund en Suecia, durante un experimento típico. Unos pocos mililitros de suero fisiológico que contienen xenón 133, isótopo radiactivo que emite rayos gamma, se inyecta en la arteria carótida de los pacientes. Una dotación de 254 detectores de centelleo colocados externamente (localizados en una caja por detrás de la cabeza del individuo) registran de esta manera la llegada y subsiguiente desaparición del isótopo radiactivo, xenón 133, de la corteza durante los dos minutos inmediatos a su inyección. Los datos de centelleo se procesan por un ordenador y se proyectan después en una pantalla de televisión en color.

afirmarse que el cerebro, consciente y en reposo, se dedica de forma preponderante a la simulación del comportamiento.

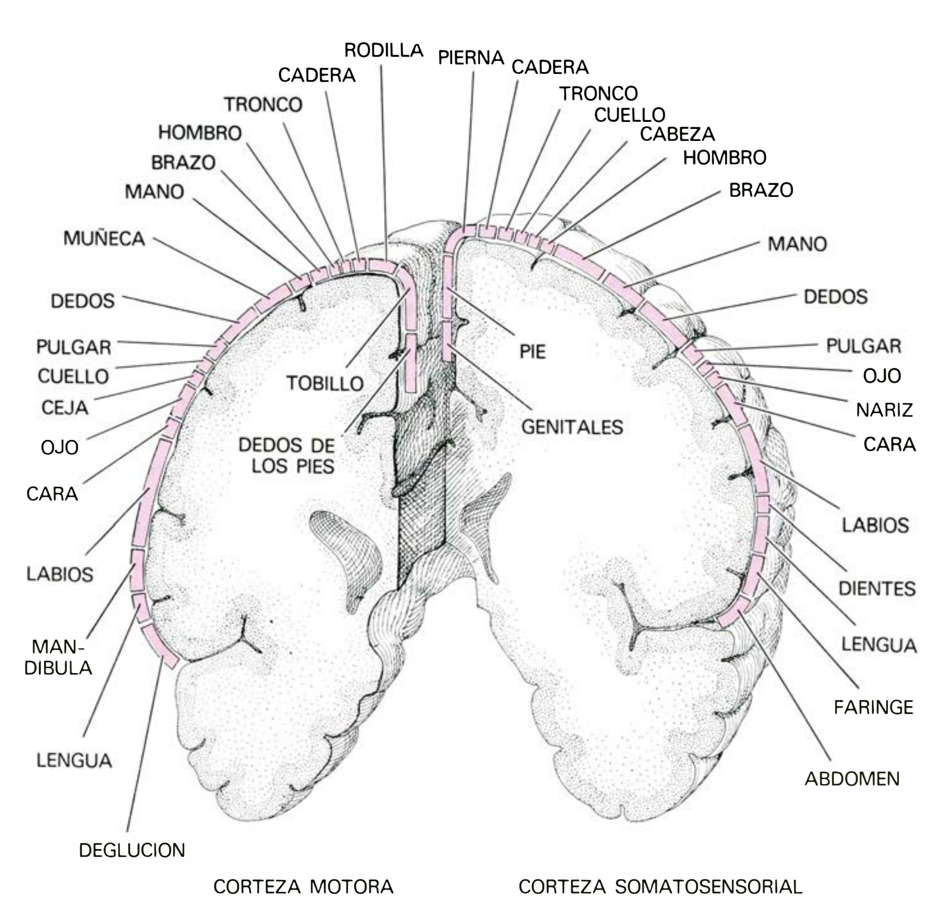
¿Cuál es el efecto que tiene un simple estímulo sensorial en la imagen típica representativa del flujo regional de la corteza? Para dilucidar este punto, programamos el ordenador de suerte que sólo se proyectaran, en color y en la pantalla, aquellas imágenes que discreparan de las obtenidas habitualmente en reposo. Cuando el individuo abría los ojos y miraba un objeto, la representación del flujo cortical experimentaba un cambio drástico: se apreciaba un aumento aproximado del 20 por ciento en el flujo correspondiente a la corteza de asociación visual, localizada en los lóbulos temporal y occipital. (No se apreciaba la corteza visual primaria, situada más profundamente en el lóbulo occipital, en la parte posterior del cerebro, porque estaba irrigada por la arteria vertebral y, por tanto, no recibía el xenón radiactivo inyectado en la arteria carótida.) Además, una parte bien definida de la corteza premotora, el campo visual frontal, se tornó activa.

La estimulación auditiva bajo la forma de un ruido intenso y sin significado específico aumentó el flujo sanguíneo en zonas próximas de la parte superoposterior del lóbulo temporal, a cada lado del cerebro, donde se localizan la corteza auditiva primaria y la corteza de asociación auditiva. El flujo en estas áreas aumentó más mediante la audición de palabras simples como "bang", "zoom" y "crack". Las regiones activadas comprenden el área de Wernicke en el hemisferio izquierdo, que está relacionado con la interpretación del lenguaje hablado. Al prestar atención a las palabras que se pronunciaban, estando los ojos cerrados, el campo visual frontal en la corteza premotora se activó ligeramente. Estímulos verbales más complejos originaron un aumento del flujo circulatorio local en la parte inferoposterior del lóbulo frontal, donde se localiza, en el lado izquierdo, el centro de Broca del lenguaje.

Los efectos de la percepción táctil fueron estudiados por Per Roland en Copenhague trabajando en colaboración con uno de nosotros (Lassen). Se pidió a las personas que indicaran verbalmente cuál de dos objetos (pequeñas barras de metal), situadas una a continuación de la otra en la



CONJUNTO DE DETECTORES colocado en la parte anterior del instrumento. Cada detector explora un centímetro cuadrado de corteza cerebral. Así, el flujo sanguíneo de todo un hemisferio se mide a la vez.



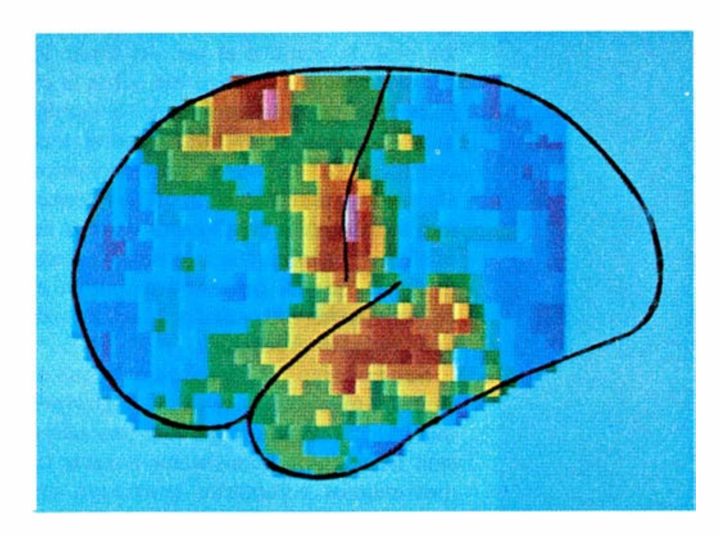
AREAS MOTORAS Y SOMATOSENSORIALES de la corteza. Constituyen las proyecciones de las áreas motoras y somatosensoriales (sensación en la piel) del cuerpo. Las áreas que representan la cara, la lengua y los dedos son desproporcionadamente grandes, debido a que la cantidad de superficie cortical dedicada a una determinada parte del cuerpo refleja los requerimientos motores y somatosensoriales de tal zona corporal.

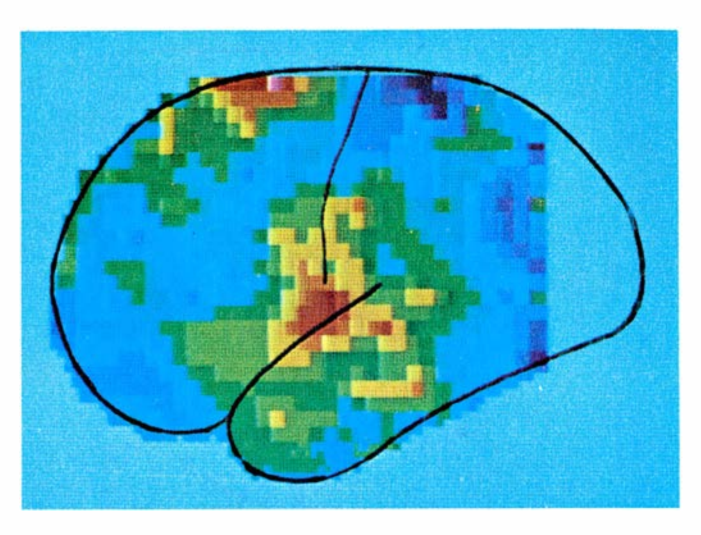
palma de la mano, manteniendo los dedos inmóviles, era la de mayor tamaño. Este estímulo táctil activó el área, correspondiente a la mano, de la corteza somatosensorial primaria de la parte central del hemisferio cerebral y del lóbulo adyacente, siempre en el lado opuesto al estimulado (como se sabe, las funciones motora y sensitiva de las extremidades están controladas por el hemisferio del lado opuesto del cuerpo).

La realización de estos sencillos tests sensoriales requieren algunas funciones de la memoria: deben retenerse los estímulos sensoriales precedentes al que se está percibiendo en un momento dado, y hay que comparar la información acumulada con la de la experiencia del momento. Se configura entonces una hipótesis sobre el significado de la entidad que se percibe; tal identificación se completa tras una búsqueda activa de los rasgos distintivos característicos. Este proceso implica asociaciones, en las cuales la memoria desempeña un papel fundamental. En nuestros experimentos constatamos que cada tipo de estímulo activa tanto la corteza sensorial primaria como el área de asociación adyacente para la modalidad sensorial específica. Estas áreas fueron las únicas activadas de forma coherente por el estímulo sensorial, lo que sugiere que en la corteza de asociación se localizan formas específicas de memoria para cada modalidad de sensación.

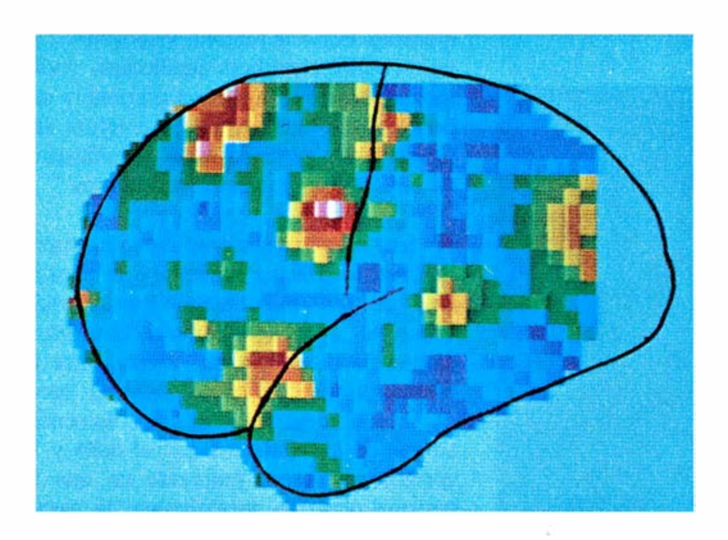
Debemos destacar que no hemos estudiado más funciones de la memoria que las relacionadas directamente con la percepción sensorial. La memoria a corto plazo (que opera, por ejemplo, en el recuerdo inmediato de un número de teléfono o de un nombre de persona) se suele considerar localizada en la región llamada hipocampo, en la parte más profunda del lóbulo temporal. Como ocurre con la corteza visual primaria, el hipocampo es irrigado a partir de la arteria vertebral, por cuya razón no aparece en nuestras gráficas basadas en la inyección intracarotídea.

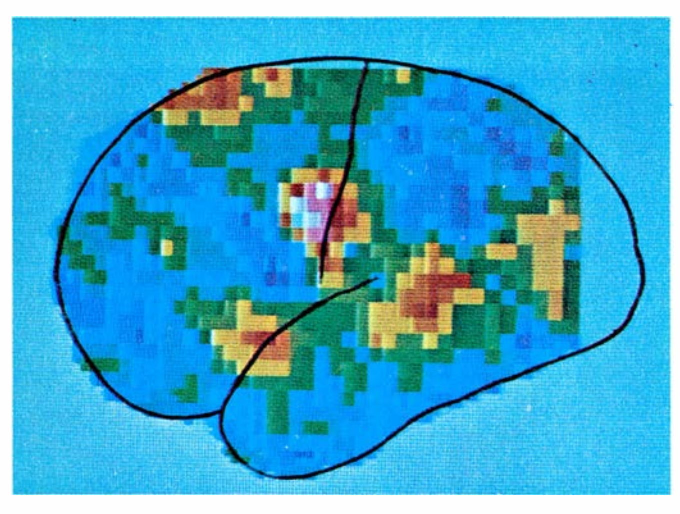
Los movimientos voluntarios de la mano fueron estudiados por Jes Olesen en Copenhague en colaboración con uno de nosotros (Lassen). En 1970, se apreció que cuando se abría y cerraba rítmicamente el puño de la parte del cuerpo opuesta a la del hemisferio cerebral que se examinaba, aumentaba el flujo circulatorio en el área de la mano de la corteza motora primaria, en la parte central del cerebro. Los movimientos de la mano determinaban también un incremento del flujo en el área adyacente a la corteza somatosensorial primaria, la cual recibe señales retroactivas procedentes de la piel, de los tendones y de los músculos de la





EL HABLA activa tres centros de cada hemisferio: el área de la boca-lengua-laringe, de la corteza motora y somatosensorial, el área motora suplementaria y la corteza auditiva. En estas imágenes pueden observarse las diferencias de activación entre los dos hemisferios que corresponden al promedio de nueve individuos diferentes: en el hemisferio derecho (*derecha*) el área de la boca-lengua-laringe está menos diferenciada y se confunden con la corteza auditiva.





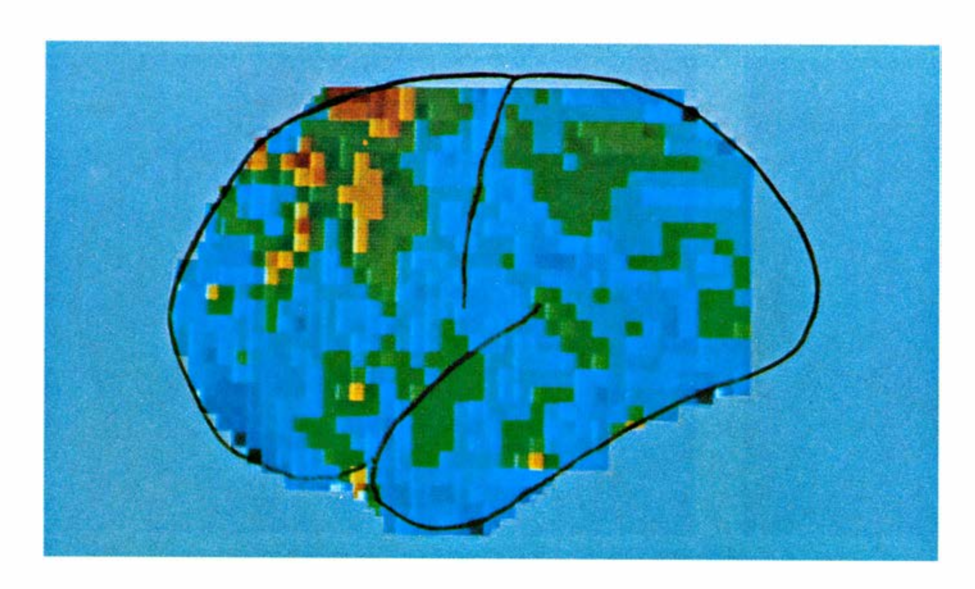
LECTURA EN SILENCIO Y LECTURA EN VOZ ALTA. Implican diferentes patrones de actividad en la corteza. La lectura en silencio (*izquierda*) activa cuatro áreas: el área de asociación visual, el campo visual frontal, el área motora suplementaria y el centro del habla de Broca, en la parte inferior del lóbulo frontal. La lectura en voz alta (*derecha*) activa otros dos centros más: el área de la boca y la corteza auditiva. Aquí se muestra el hemisferio izquierdo para ambos casos, aunque se han obtenido resultados similares en el hemisferio derecho. Considerando el isótopo radiactivo, el acto de la lectura en voz alta requiere la actividad simultánea de siete centros corticales discretos, los cuales se hallan dispuestos formando una figura en forma de *Z* en cada hemisferio.

mano mientras ésta se mueve. Sin embargo, como podía esperarse, cuando la mano que se abría y se cerraba era la correspondiente al mismo lado del cuerpo que el hemisferio explorado, no se producía ningún cambio en el flujo habitual.

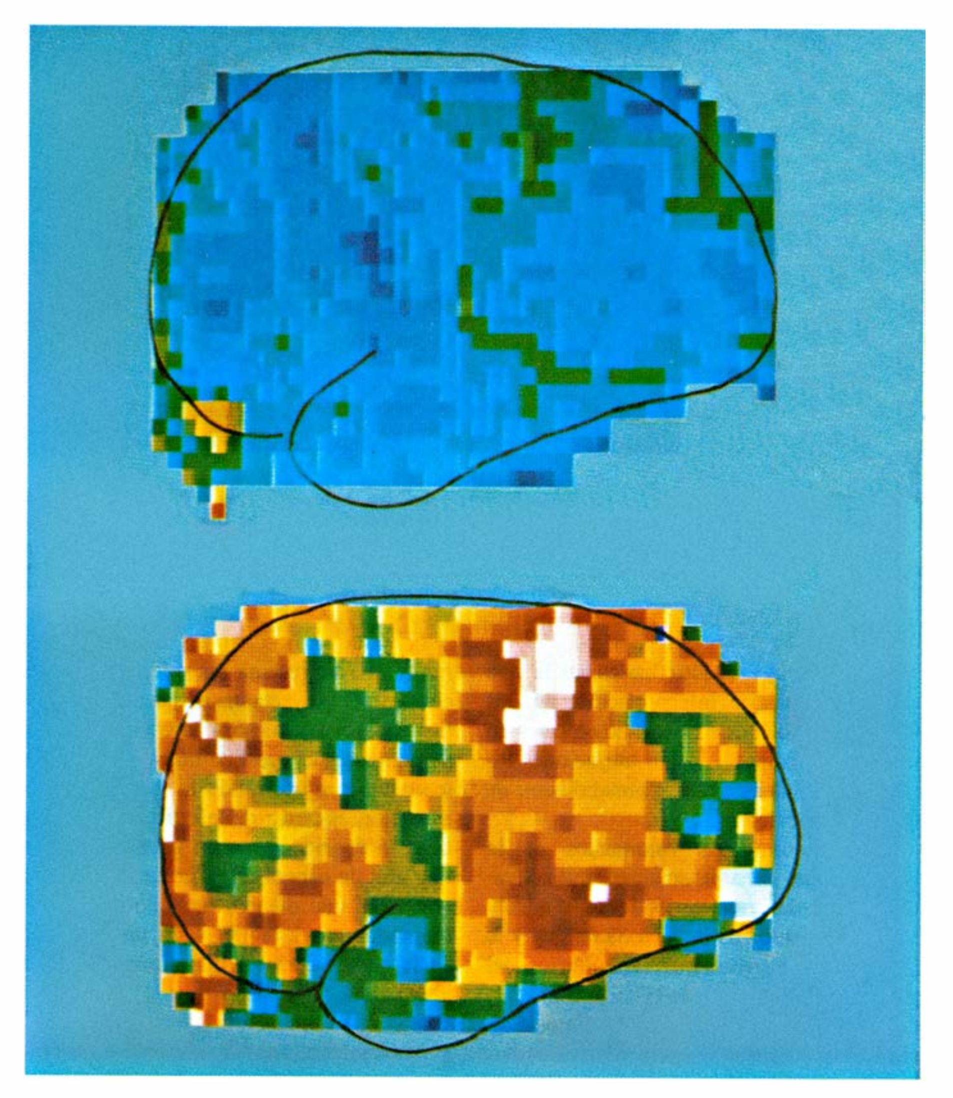
Recientemente, Marcus E. Raichle y sus colaboradores, del Mallinckrodt Institute of Radiology en San Luis, estudiaron el control de los movimientos de la mano inyectando un isótopo radiactivo del oxígeno en las arterias que conducen la sangre hacia el cerebro, al objeto de controlar de forma directa el consumo local de oxígeno por el tejido cortical. Encontraron que los movimientos de la mano aumentaban la utilización de oxígeno en las mismas partes de la corteza en las que había observado un aumento en el flujo regional. Este hecho constituye una prueba directa que apoya la idea básica en la que descansa nuestra interpretación de los datos de flujo: que los cambios locales en el flujo sanguíneo reflejan variaciones locales en la intensidad del metabolismo de la célula nerviosa.

Los movimientos voluntarios de la boca durante el habla originan una activación clara del área cortical que controla los movimientos de la boca, la lengua y la laringe; los movimientos voluntarios del pie estimulan la parte de la corteza motora del hemisferio opuesto situada en posición superior a la parte activada por los movimientos de la mano. Estos y otros hallazgos confirman que las áreas somatosensorial primaria y el área motora se hallan organizadas en dos bandas estrechas y adyacentes, que se extienden de oído a oído a través de la parte superior de la corteza. El mapa que representa las diferentes partes del cuerpo proyectadas en las áreas somática sensorial y motora se conoce detalladamente desde que Penfield y sus colegas lo elaboraron utilizando la técnica de estimulación eléctrica de la corteza. El mapa tiene la apariencia de un homúnculo con una amplia cabeza dirigida hacia el lóbulo temporal, una mano y un dedo pulgar aumentados en el centro y un pie reducido en la parte superior, que alcanza la zona interna del hemisferio.

Las contracciones de la musculatura voluntaria activan también la corteza premotora, localizada en la parte superior del lóbulo frontal. Esta activación tiene lugar en los dos hemisferios cerebrales y, por lo que sabemos, se localizan siempre en la misma área, independientemente de



"HABLA INTERNA" bajo la forma del contaje repetido de manera silenciosa desde 1 a 20; activa las regiones frontales, en particular el área motora suplementaria. Este tipo de habla "imaginado" no activa de modo significativo las áreas que normalmente están implicadas en el lenguaje, tales como las áreas de Broca y de Wernicke. La imagen representa pues la localización de un acontecimiento puramente mental.



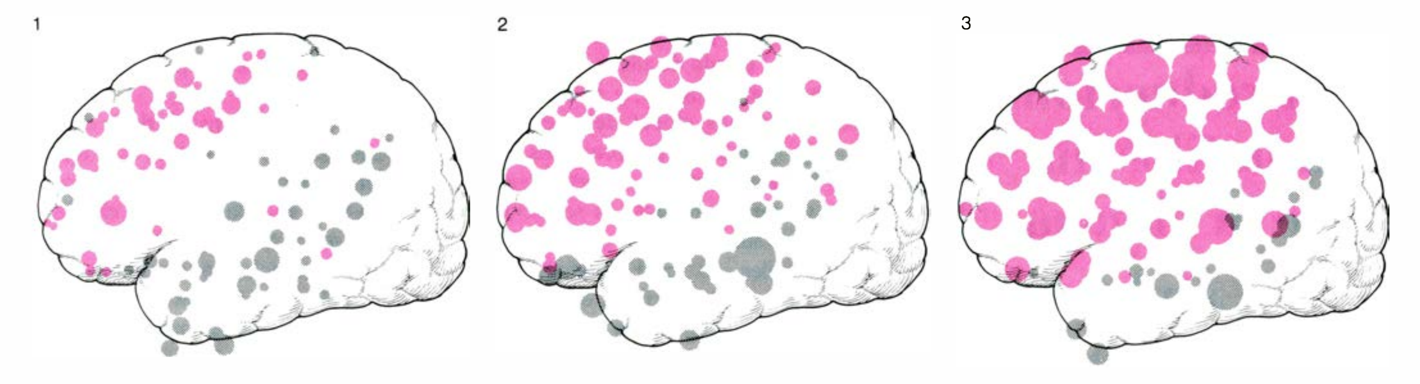
ACTIVACION GLOBAL de la corteza cerebral; se observa cuando el individuo está acometiendo intensa o emocionalmente una tarea. La parte superior de la figura corresponde a la imagen del hemisferio izquierdo de un hombre anciano; el flujo medio global, que es de 33 mililitros por 100 gramos de tejido cerebral por minuto, es bajo. Durante el proceso representado por la imagen inferior, se ordenó al individuo que hiciera movimientos rítmicos de apertura y cierre de su mano derecha. El flujo medio global se incrementó hasta 48 mililitros por 100 gr. y por min. aumentos en el área de la mano y en la corteza auditiva.

que la parte del cuerpo que se mueva sean los ojos, la mano o el pie. Al situar los detectores de nuestro aparato sobre la cabeza de la persona examinada, el cambio más destacado en el flujo cerebral local ocurre cerca del surco anteroposterior afectando a la región de la corteza premotora situada en la cara interna de cada hemisferio y denominada área motora suplementaria. Se sabe que este área influye en las actividades motoras complejas de todo tipo, incluido el lenguaje.

Sin embargo, la apreciación de que en el área suplementaria y en la corteza premotora circundante se produce una activación tan masiva y consistente durante los movimientos voluntarios constituyó una sorpresa para nosotros. Comprobamos que la estimulación de las áreas motoras suplementarias era más pronunciada durante la acción dinámica muscular, como la que representa el escribir a máquina, que durante las contracciones musculares estáticas. Por este motivo, y apoyados por las pruebas científicas publicadas, podemos afirmar que la corteza premotora superior, incluyendo el área premotora suplementaria, participa en la planificación de las acciones motoras secuenciales.

Tiene interés mencionar aquí un experimento reciente referente a la naturaleza de los movimientos voluntarios. Tanto en Copenhague como en Lund estudiamos la diferencia que media entre la representación del flujo cerebral local durante el desarrollo de una secuencia simple de movimientos con el dedo y la correspondiente al momento en que la persona planeaba mentalmente la ejecución de la citada sucesión de movimientos. Dándole instrucciones adecuadas, se lograba que el individuo imaginara, mentalmente, todos los movimientos de la secuencia con una ordenación temporal correcta, pero manteniendo de hecho en todo momento el dedo absolutamente inmóvil; el movimiento ficticio activaba el área motora suplementaria. Cuando se ejecutaba realmente toda la serie de movimientos, se estimulaban también el área de proyección de los dedos de la mano en la corteza motora primaria y las áreas correspondientes de la corteza sensorial. De estos hallazgos se desprende que el área motora actúa a modo de programador de los movimientos, la corteza sensorial primaria realiza funciones de controlador y la corteza motora primaria de ejecutor.

Asimismo, hemos investigado con detalle los procesos que se desarrollan durante el lenguaje oral. Nos llamó la atención que tanto el hemisferio derecho como el izquierdo mostraran una actividad en gran manera equiparable. Como quedó expuesto más arriba, la audición de palabras escuetas activa la corteza auditiva de



ESTADO DE REPOSO, **sensación táctil y dolor moderado: parecen constituir tres niveles progresivos de consciencia. Se sometió a estudio ocho individuos (*1*) en estado de reposo, (*2*) durante la estimulación eléctrica umbral del pulgar y (*3*) durante una estimulación con intensidad correspondiente a 2,5 veces la umbral, que originó una sensación moderada de dolor. A lo largo de esta secuencia, los niveles medios de flujo sanguíneo cerebral y de consumo de oxígeno aumentaron progresivamente. Los círculos coloreados indican intensidades de flujo superiores a la media y, los círculos grises, inferiores a ella.**

los dos hemisferios. El hablar en voz alta estimula tres áreas más, que son las correspondientes a las proyecciones de la cara, de la lengua y de la boca en las cortezas somatosensoriales y motora, la corteza premotora superior de ambos hemisferios (que incluye el área motora suplementaria) y el área de Broca en la parte inferoposterior del lóbulo frontal izquierdo, así como tambien la parte correspondiente a ésta en el lóbulo frontal derecho. Durante la lectura en voz alta se produce, además, la activación de la corteza de asociación visual y de los campos visuales frontales y la activación de la corteza visual primaria (esto lo inferimos sin haberlo podido confirmar experimentalmente, debido a que no se logra visualizar esa zona con nuestra técnica). Así pues, siete regiones discretas, o delimitadas, de la corteza muestran una actividad simúltanea; las cuales se hallan dispuestas formando una figura en forma de *Z* en la superficie de cada hemisferio. Resultó interesante comprobar la diferencia existente entre la lectura en voz alta y la lectura en silencio; en este último caso no se produjo la activación de las áreas de la boca de las cortezas somatosensorial y motora ni de las áreas auditivas, pero sí denotaron actividad las cinco áreas restantes.

Los estudios llevados a cabo sobre los efectos que tienen las lesiones cerebrales en el lenguaje han demostrado que la destrucción del área de Broca del hemisferio izquierdo produce afasia motora, es decir, pérdida de la capacidad del habla coordinada. La destrucción del área correspondiente a la citada en el hemisferio derecho no conlleva, sin embargo, ninguna alteración discernible en el lenguaje. Por ello nos sorprendió que esta parte del hemisferio derecho se mostrara activa durante la expresión verbal, pues sugería que debía contribuir en alguna forma (aunque no esencial) a la síntesis final y ejecución del habla. En Copenhague, Borge Larsen, en colaboración con uno de nosotros (Lassen), profundizó en la investigación que realizaba tratando de encontrar alguna diferencia entre las respuestas de flujo circulatorio inducidas por el lenguaje en cada hemisferio cerebral. Aunque por razones de ética no pudo determinar el flujo en el hemisferio derecho e izquierdo del mismo individuo, los resultados obtenidos por Larsen sugieren la existencia de las siguientes diferencias: en el hemisferio izquierdo se producen generalmente aumentos de flujo por separado en el área de Broca y en la corteza auditiva, mientras que en el hemisferio derecho es frecuente que ambas zonas formen una región activa confluyente. Además, el área motora suplementaria del hemisferio izquierdo muestra habitualmente mayor actividad durante el lenguaje que la correspondiente del hemisferio derecho.

El análisis de la estimulación cortical durante la lectura indica que se está desarrollando una actividad compleja por parte de varias regiones circunscritas de la corteza que intervienen de manera específica. Este sistema equivale a un programa de ordenador en el que se utilizaran diferentes subrutinas en razón del problema a resolver. En general, nuestros resultados confirman las conclusiones a las que llegó A. R. Luria, de la Universidad estatal de Moscú, basándose en sus análisis neuropsicológicos llevados a cabo en pacientes con lesiones cerebrales: "Los procesos de comportamiento complejo no se encuentran localizados sino distribuidos en el cerebro, y la contribución de cada zona cortical al conjunto del sistema funcional es altamente específica".

En nuestros estudios sobre diversas formas de activación del cerebro quedó puesto de manifiesto pronto que muchos de nuestros pacientes en estado de vigilia no sólo mostraban aumentos locales en el flujo sanguíneo de regiones específicas de la corteza sino también una elevación global del flujo cerebral. Jarl Risberg y uno de nosotros (Ingvar) encontraron que los tests psicológicos basados en pruebas memorísticas y deductivas sencillas originaban, además de los cambios locales, un aumento total del flujo cerebral de un 10 por ciento. Esta elevación integral del flujo sanguíneo parece estar claramente relacionada con el esfuerzo desarrollado por el individuo en la realización del test: no se produce cuando se llevan a cabo pruebas fáciles, pero es evidente cuando la persona se esfuerza al enfrentarse a un problema difícil.

Estos hechos apoyan la distinción, señalada por los neurofisiólogos, entre las vías específicas y las no específicas del cerebro. Las vías sensoriales y motoras específicas llegan y surgen de zonas bien definidas de la corteza, que entran en actividad durante la recepción e interpretación de mensajes sensoriales concretos y durante la ejecución de una función motora determinada. Las funciones que exigen cierto esfuerzo activan al mismo tiempo áreas corticales más amplias a través de vías difusas que se extienden a partir del sistema reticular en el tronco encefálico y del tálamo en el mesencéfalo. Los estudios llevados a cabo en animales muestran que, sin la estimulación de las vías no específicas difusas, las vías específicas no alteran de forma apreciable por sí mismas la actividad del cerebro fuera de las regiones donde se proyectan. Por tanto, todo indica que, para que el cerebro "comprenda" el mundo que le rodea, perciba su significado y actúe en el desarrollo de funciones difíciles, debe activarse la corteza cerebral no sólo de forma localizada sino también en su totalidad.

Las investigaciones sobre la ansiedad y el dolor sustentan firmemente que, durante la realización de distintas funciones, existe una correspondencia entre la activación generalizada de la corteza y el grado de consciencia. La amenaza que para el cuerpo o para el bienestar psíquico representan un estímulo de dolor o un fuerte

estado de ansiedad provoca una gran intensificación del estado de alerta hacia uno mismo y hacia el ambiente, y suscita asimismo un aumento global del metabolismo y flujo sanguíneo cerebral. Los estados de ansiedad activan diversas vías cerebrales y también inducen la secreción de la hormona adrenalina por las glándulas suprarrenales, la cual produce un efecto generalizado de alerta en el cuerpo. Bo Siesjö y sus colaboradores, del Hospital de la Universidad de Lund, han realizado estudios recientes en ratas mientras éstas se iban recuperando del estado de anestesia tras haber sido paralizadas con una droga curarizante. La ansiedad de los animales despertándose en un estado de parálisis dio lugar a un gran aumento en la intensidad de consumo de oxígeno cerebral y del flujo sanguíneo; este efecto podía, sin embargo, aminorarse mediante la extirpación de las glándulas suprarrenales y podía ser totalmente suprimido administrando drogas bloqueantes de los receptores de la noradrenalina.

Recientemente estudiamos en Lund el efecto que produce un estímulo doloroso (una descarga eléctrica aplicada sobre la piel) en la imagen normal del flujo sanguíneo cerebral. (Todos estos experimentos fueron, por supuesto, llevados a cabo con el consentimiento de los pacientes, plenamente informados, y como base para las pruebas de diagnóstico clínico.) Cuando la intensidad del estímulo era baja, justo por encima de la del umbral de percepción, la descarga producía una sensación aparentemente táctil; cuando esta intensidad era elevada, la descarga producía una sensación de dolor moderado. Como podía esperarse, el flujo cerebral no alteró de forma perceptible el flujo medio del hemisferio cerebral ni tampoco el consumo de oxígeno; produjo, sin embargo, un pequeño aumento en el flujo local de las partes anterior y superior del lóbulo frontal. Por su lado, el estímulo de dolor moderado dio lugar a una elevación global del 20 por ciento del flujo medio del hemisferio y del consumo de oxígeno, y se produjeron aumentos sobre el valor medio locales en los lóbulos frontales. Por tanto, el estímulo táctil y el de dolor moderado parecen inducir un aumento de la atención o de la consciencia del cerebro.

Estas observaciones sustentan la hipótesis de que la actividad generalizada del cerebro conlleva, no sólo una acentuación de la actividad electroencefalográfica, sino también un aumento del flujo cerebral y del consumo de oxígeno, y que esta reacción se halla relacionada con una intensificación del grado de consciencia. La activación global del cerebro se relaciona también, probablemente, con los componentes emocionales de la experiencia, aunque no hemos realizado todavía observaciones sistemáticas en dicho sentido.

¿Cómo serán los mapas de la corteza cerebral en el futuro? La técnica del xenón radiactivo (y una modificación incruenta de la misma en la que el isótopo radiactivo se inhala en vez de inyectarse) es algo rudimentaria. Los resultados reflejan principalmente lo que ocurre en las capas superficiales de la corteza y dejan de lado lo que sucede en las estructuras más profundas; además la resolución temporal es baja (como máximo minutos). Estas limitaciones, unidas a recientes avances técnicos, han inspirado perfeccionamientos de nuestra técnica. Se están desarrollando instrumentos multidetectores complejos para determinar la distribución tridimensional del radioisótopo en la cabeza, al objeto de que pueda analizarse la actividad de las partes más profundas del cerebro.

Además, la aplicación de isótopos radiactivos a la investigación de la función cerebral ha abierto el nuevo campo de los estudios metabólicos regionales. Louis Sokoloff y Martin Reivich, del National Institute of Mental Health en los Estados Unidos, han estudiado el metabolismo cerebral a escala microscópica inyectando en el cerebro un análogo radiactivo de la glucosa; la tasa de utilización de la substancia por las células nerviosas constituye un reflejo de la actividad funcional de éstas. Estos experimentos han puesto de manifiesto que la intensidad metabólica varía según normas coherentes en regiones muy pequeñas del cerebro durante la realización de diversas actividades. Por ejemplo, se ha comprobado que, en los monos, la iluminación de un ojo altera de forma discernible el consumo de glucosa por columnas de células nerviosas de la corteza cerebral separadas entre sí menos de un milímetro; estas unidades de columnas funcionales corresponden a las mostradas mediante el empleo de técnicas neurofisiológicas por David H. Hubel y Torsten N. Wiesel, de la Facultad de Medicina de Harvard. La técnica de Sokoloff-Reivich está sufriendo ciertas modificaciones con vistas a su aplicación clínica.

La generación venidera de instrumentación y de técnicas integrativas de gran potencia servirá de base para el establecimiento de un nuevo tipo de neurofisiología clínica: una era en la que los estados circulatorios y bioquímicos que acompañan a las funciones del cerebro humano podrán ser localizados con precisión y determinados cuantitativamente. Estos métodos nos llevarán a una perfección más fiel de las intrincadas normas de actividad sobre las que descansa el funcionalismo del más complejo de todos los sistemas biológicos, el cerebro humano.

La partícula hypsilón

Su inesperado descubrimiento como la partícula más pesada ha impulsado a los físicos a introducir un quark de masa mayor, elevando de cuatro a cinco el número de estas subpartículas elementales aún no observadas

Leon M. Lederman

La búsqueda de los últimos constituyentes indivisibles de la materia, que empezó con los filósofos atomistas presocráticos, continúa sin cesar 2400 años después. En las últimas décadas, el número de partículas subatómicas identificadas se han elevado a más de 100, al desarrollarse máquinas muy potentes capaces de hacer chocar con violencia trozos de materia y estudiar los productos de la colisión. Al principio, los físicos creían que estas partículas no podían romperse en entidades más pequeñas. Después encontraron que sólo los cuatro leptones (el electrón, el muon y las dos clases de neutrinos) parecían ser realmente elementales en el sentido de carecer de un tamaño que pudiera medirse y de partes constituyentes. Las restantes partículas, los hadrones (que comprenden el protón, el neutrón y el pion), han resultado ser objetos complejos que muestran claros signos de poseer una estructura interna. En 1964, la hipótesis de los quarks que, desde entonces, ha venido constituyendo la piedra fundamental de la física de partículas, se introdujo como una descripción de esta estructura. Se suponía que todos los hadrones podían explicarse en términos de tan sólo tres entidades elementales llamadas quarks. Razones teóricas y pruebas experimentales postularon pronto la existencia de un nuevo quark. Aunque no se ha podido observar ninguno de los cuatro quarks, a pesar de los múltiples intentos por aislar uno, hay claros motivos para creer en su existencia.

El año pasado, un grupo de investigadores (del que yo formaba parte) de la Columbia University, la State University de Nueva York en Stony Brook y el Fermi National Accelerator Laboratory (Fermilab) descubrieron una nueva partícula con una masa equivalente a una energía de 9,4 GeV (1 GeV es mil millones de electronvolts), masa ésta que triplica, y pasa, la de cualquier otra entidad subatómica identificada hasta entonces. Se la denomina partícula hypsilón (Υ) y sugiere la existencia de un quinto quark, cuya masa sea mayor que cualquiera de los restantes. Ya que los cuatro quarks originales daban cuenta de todas las propiedades conocidas de los hadrones, una quinta subpartícula parece superflua. Se diría que su existencia aporta un dudoso beneficio a la hipótesis de los quarks. Por una parte debería ayudar a los físicos a determinar la naturaleza, aún no aclarada, de las fuerzas que caracterizan a los quarks. Por otra, una verdadera proliferación de los mismos podría hacer tambalear la hipótesis central del modelo que los considera como los últimos constituyentes de la materia. Primero se introdujeron los quarks para explicar el número, en continuo aumento, de hadrones. Ahora son los quarks los que crecen en número, y no parece existir ninguna razón teórica que no permita descubrir la existencia de otros más pesados incluso.

La investigación que ha conducido al descubrimiento de la partícula hypsilón comenzó en 1967 en el Brookhaven National Laboratory. Mediante el sincrotrón de 30 GeV de Brookhaven se lanzaban protones (p) muy energéticos sobre núcleos de uranio, que están formados por protones y neutrones; a éstos se les llama colectivamente nucleones (N). Deseábamos saber lo que sucedía cuando se producían un par de leptones de carga opuesta (l^- y l^+), una reacción que podía escribirse como $p + \mathrm{N} \rightarrow l^- + l^+ +$ otras cosas; donde "otras cosas" significaba que no estábamos interesados en las otras partículas que se producían. Antes de pasar a describir los experimentos realizados, digamos algo más sobre los leptones para que el lector pueda entender mejor por qué razón hemos trabajado intensamente en ellos a lo largo de 10 años.

Los leptones se distinguen de las demás partículas subatómicas en que no se hallan sometidos a las fuerzas "fuertes" que ligan a los protones y neutrones para formar los núcleos atómicos. A consecuencia de ello, los leptones de gran energía tienen un vigoroso poder de penetración en la materia. Uno de ellos, el neutrino (ν), que carece de carga eléctrica, podría atravesar millones de kilómetros de plomo sin chocar con nada. El muon (μ), que es unas 200 veces más pesado que el electrón (e^-), pero que por lo demás tiene idénticas propiedades, es frenado cuando se mueve a través de la materia, ya que debe arrastrar una carga eléctrica a través de otras cargas eléctricas. Sin embargo, como tales fuerzas electromagnéticas son 100 veces más débiles que las fuerzas fuertes, el muon puede penetrar muchos metros de hierro. Con una carga eléctrica igual que la del muon, el electrón se detiene con mayor facilidad debido a su menor masa; no puede abrirse camino a través del hierro como lo hacen los leptones más pesados.

El par leptónico ($l^- + l^+$), creado en la reacción que estudiamos en Brookhaven, tenía las mismas propiedades cuánticas que el cuanto de energía electromagnética: el fotón (γ). Ello era obvio, si atendemos a la facilidad con que un fotón se transforma en un par muónico ($\mu^- + \mu^+$), o en un par electrón-positrón ($e^- + e^+$), reacciones que se simbolizan en el formalismo: $\gamma \rightarrow \mu^- + \mu^+$ y $\gamma \rightarrow e^- + e^+$.

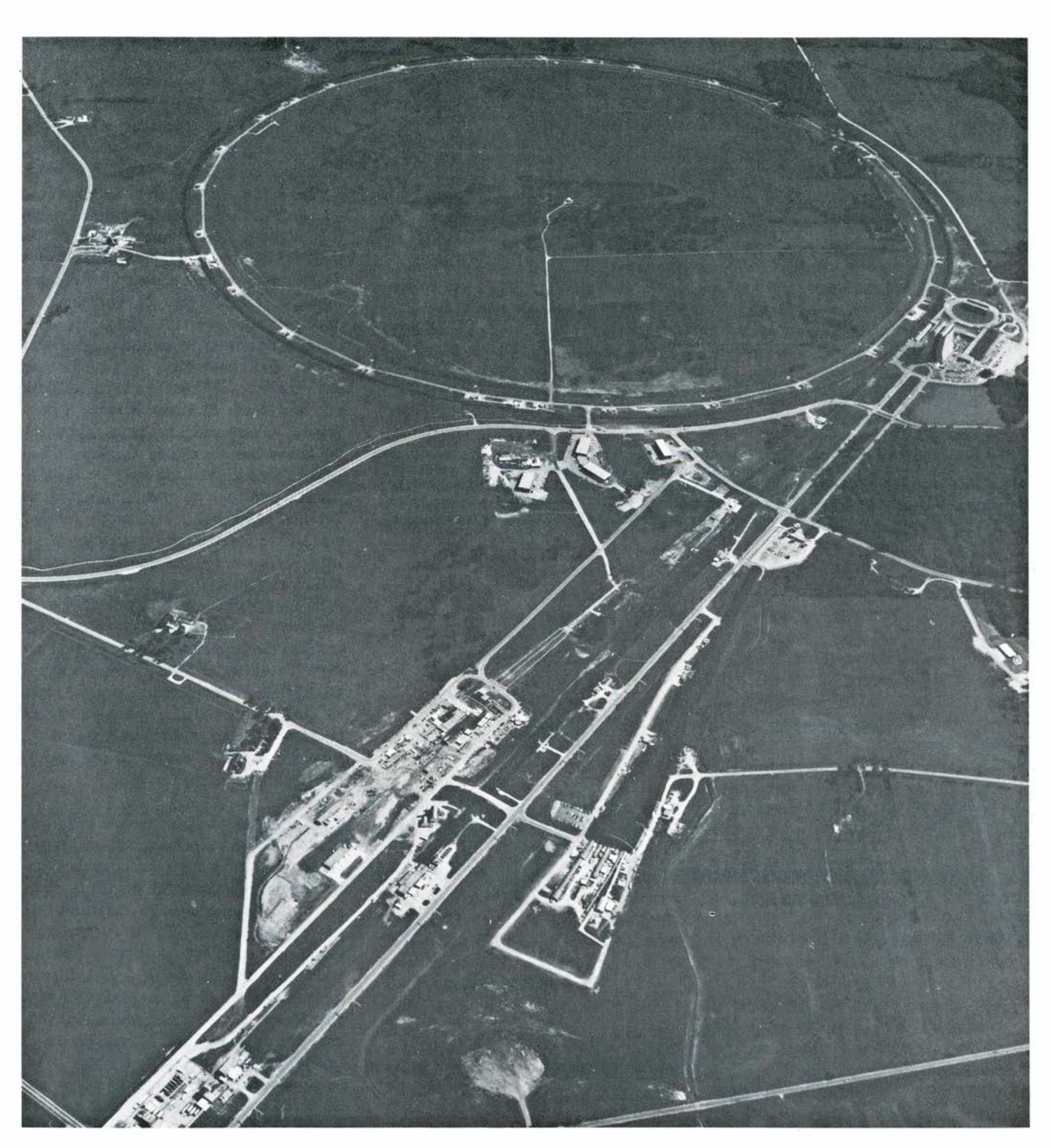
La masa constituye una importante razón de diferencia entre los fotones y los pares leptónicos. Mientras que un par leptónico tiene una masa en reposo positiva cuando se le considera como una única partícula que se mueve con una velocidad igual a la suma vectorial de los movimientos de sus dos componentes, el fotón posee siempre masa en reposo nula. Para inter-

NOTA DEL EDITOR

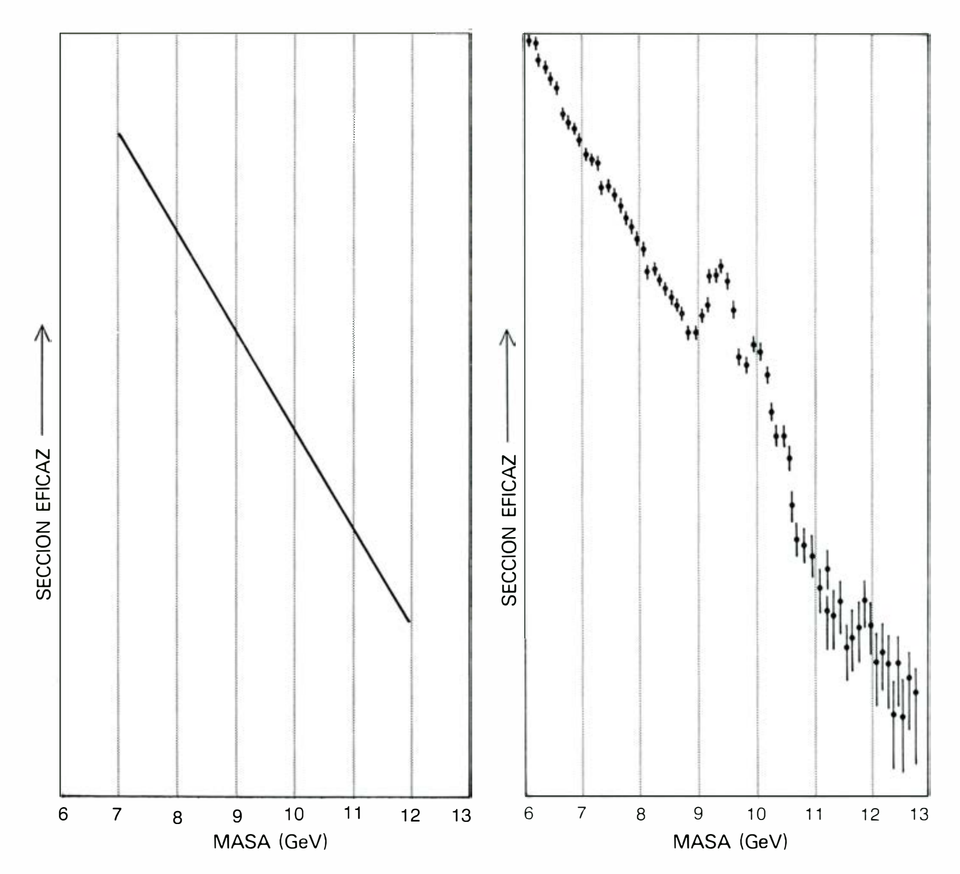
El autor desea agradecer la contribución al trabajo descrito en este artículo aportada por: J. A. Appel, B. C. Brown, C. N. Brown, J. H. Christenson, S. W. Herb, G. S. Hicks, D. C. Hom, W. R. Innes, A. S. Ito, H. Jöstlein, D. M. Kaplan, R. D. Kephart, P. J. Limon, B. G. Pope, J. C. Sens, H. D. Snyder, K. Ueno, T. Yamanouchi, J. K. Yoh y E. Zavattini.

pretar tal diferencia podemos tratar el par leptónico como siendo el producto de la desintegración de un pseudo-fotón de vida corta llamado fotón virtual. La noción de fotón virtual aparece también en otras reacciones en las que se examinan las propiedades eléctricas y magnéticas de la materia. Las leyes de conservación de la energía y el momento nos permiten calcular de forma totalmente rutinaria la masa, energía y momento del fotón virtual primitivo, a pesar de su naturaleza efímera. Para determinar su masa (M) debemos medir la energía de las partículas l^- y l^+ que se producen en la colisión. La masa viene dada por la fórmula $M^2 = 4E^-E^+ (sen^2\theta/2)$ donde θ es el ángulo de salida de un leptón con relación al otro y E^-E^+ el producto de sus energías. Se ha supuesto que cada una de estas energías es grande comparada con la masa del leptón.

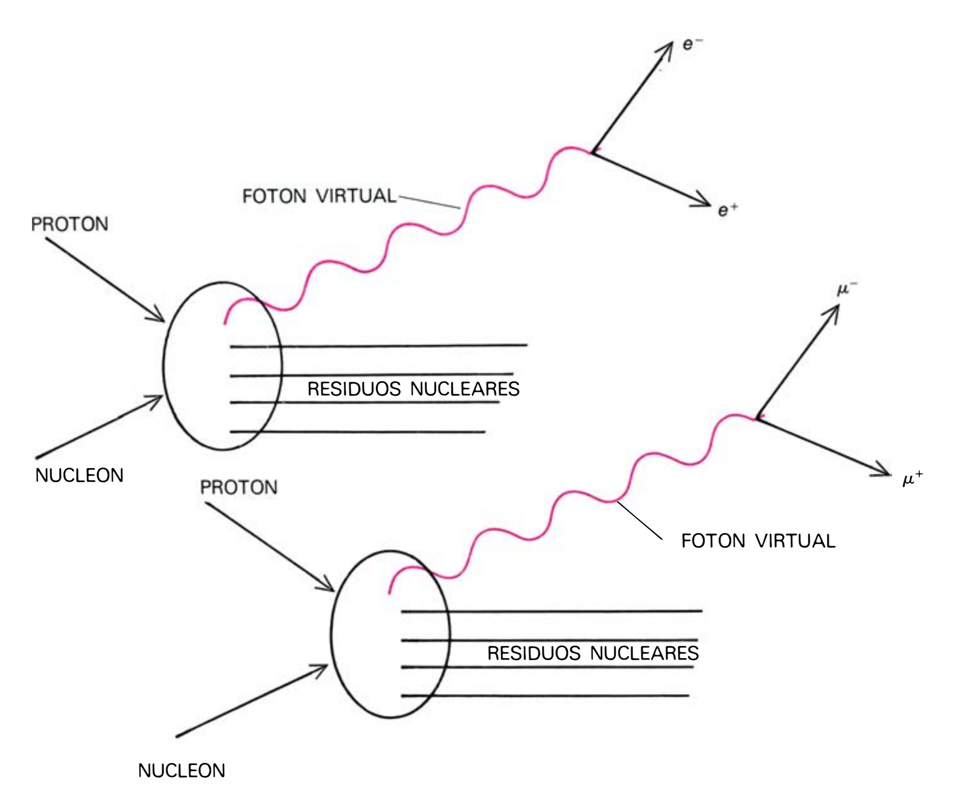
Ya en 1967 nos dimos cuenta, de una forma vaga e intuitiva, que la emisión de fotones virtuales podría remitir a domi-



SINCROTON DE PROTONES del Fermi National Accelerator Laboratory (Fermilab); se utilizó para generar los pares muónicos en los experimentos que condujeron al descubrimiento de la partícula hypsilón. Aquí, la máquina de 400 GeV (400.000 millones de electronvolts) aparece como el gran círculo, cuya circunferencia mide unos seis kilómetros. Tangentes al círculo se disponen largos túneles que llevan las partículas a las estaciones experimentales. Este trabajo sobre la partícula hypsilón se realizó en el laboratorio de protones, que es la extensa zona sita en el extremo inferior izquierdo. El Fermilab ocupa una superficie de 6 × 7,5 kilómetros a unos 45 kilómetros al sudoeste de Chicago en Batavia, Illinois. Fue construido por la Atomic Energy Commission mediante un contrato con un consorcio de 53 universidades. Las instalaciones del laboratorio son usadas por grupos de todo el mundo. Bajo la dirección de Robert R. Wilson, el acelerador que empezó en 1972 con una energía de 200 GeV, se mejoró hasta alcanzar los 300 GeV en 1973 y los 400 GeV en 1974.



MASAS DE LOS FOTONES VIRTUALES, que se desintegraban en pares muónicos. Se esperaba, para estas masas, una distribución continua (*izquierda*). Así sucedió, pero aparecía además una protuberancia inesperada alrededor de 9,4 GeV (*derecha*). Tal protuberancia, llamada resonancia, marcaba la presencia de la partícula hypsilón. Las barras verticales de error, que aparecen en cada dato, corresponden a la indeterminación de donde debía representarse. Cuanto menor es el número de sucesos, mayor la indeterminación.

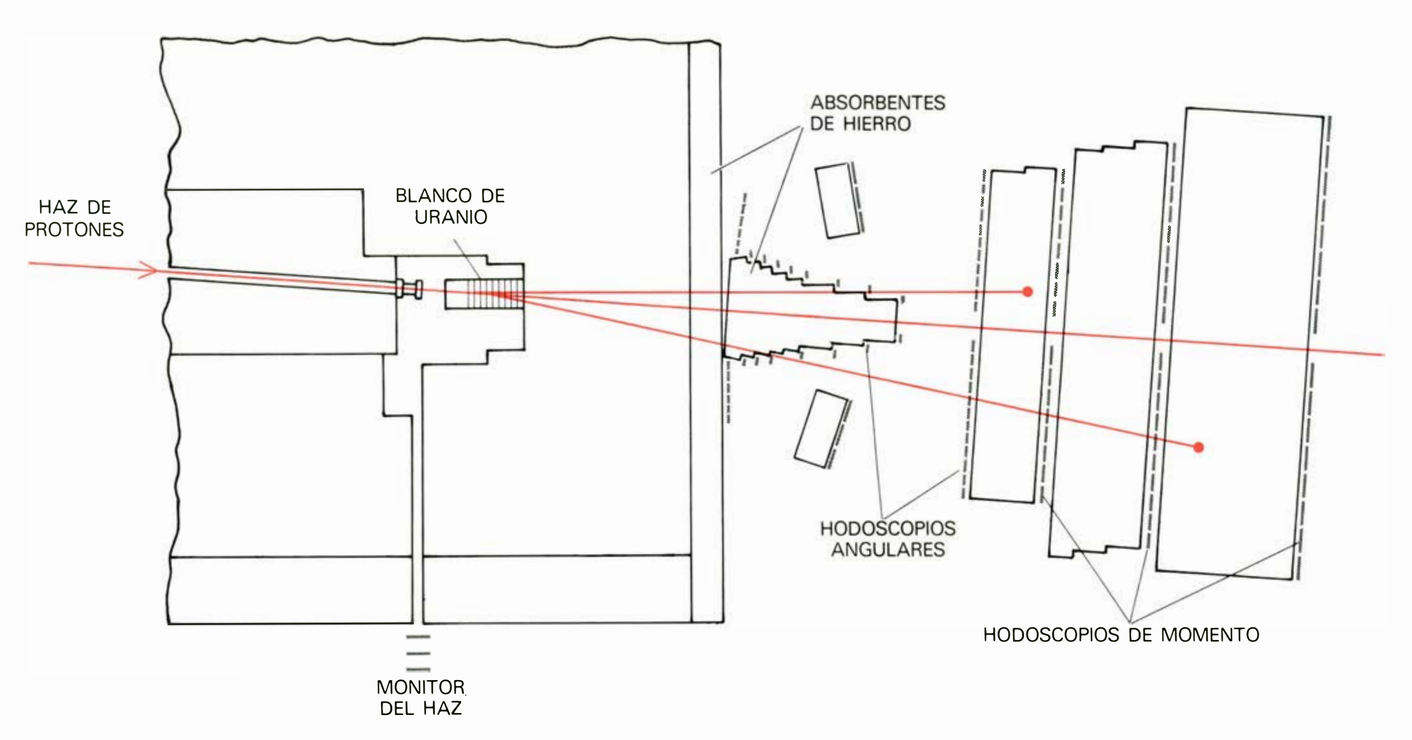


LAS COLISIONES ENTRE PROTONES Y NUCLEONES (protones o neutrones) generan algunas veces fotones virtuales que decaen inmediaramente en pares electrón-positrón (e^-e^+) o en pares de muones de carga opuesta ($\mu^-\mu^+$). La reacción inferior exige mucha más energía que la de arriba, pues el muon tiene una masa 200 veces mayor que la del electrón. La masa adicional del muon hace que éste pueda penetrar mucho más profundamente en la materia. Los residuos nucleares tendrán muy poco poder de penetración.

nios inexplorados en el interior de las partículas nucleares que chocaban. Nuestro razonamiento era como sigue: cuando un protón extraordinariamente energético chocase con un blanco formado por un nucleón, se generaría un estado excitado y complejo. La mayoría de las veces tal estado perdería energía al emitir partículas de interacción fuerte como son los piones y los kaones: Pero en otras la desexcitación se debería, en parte, a la emisión de fotones virtuales que se desintegrarían inmediatamente en pares leptónicos.

Esperábamos que las masas de las partículas virtuales, calculadas a partir de mediciones efectuadas sobre los leptones, estarían distribuidas sin solución de continuidad [*véase la ilustración superior a la izquierda*]. Como nos habíamos dado cuenta que sería más fácil producir masas menores que masas grandes, pensábamos que la producción de fotones virtuales disminuiría rápidamente al aumentar la masa. Aunque no esperábamos que los cálculos de masas se centraran alrededor de un valor particular, deseábamos que eso sucediera. Tal concentración se llama una resonancia. Si se manifestaba la resonancia, indicaría que los pares leptónicos no emanaban de alguna entidad virtual sino de alguna partícula real. De acuerdo con el principio de indeterminación de Werner Heisenberg, podríamos estimar entonces el tamaño del material (en el interior de los nucleones que han chocado) que ha servido de fuente de la nueva partícula. El principio de indeterminación de Heisenberg sugiere que cuanto mayor sea la masa de la partícula menor es el tamaño de su fuente. Esto significaba que si descubríamos resonancias de una masa suficientemente elevada, estaríamos detectando en realidad estructuras extraordinariamente pequeñas dentro de los nucleones empleados como blanco.

En 1967, iniciamos nuestra búsqueda de tales estructuras dentro de los nucleones blanco, a pesar de la creencia generalizada entonces según la cual la materia en estados altamente excitados era suave y homogénea. Y, aun cuando existieran estas entidades pequeñas y masivas, podría suceder que nuestro equipo experimental no fuera lo bastante sensible para detectarlas. Otros experimentadores habían descubierto ya que las resonancias de masa baja eran extraordinariamente raras; en el acelerador de Brookhaven, un par leptónico con una masa cercana a la del protón se crearía sólo una vez en un millón de colisiones. Masas mayores se producirían a intervalos todavía más distanciados, y por cada una que se produjera se crearían también millones de partículas que interaccionaran fuertemente. Nuestro detector tendría que poder separar los es-

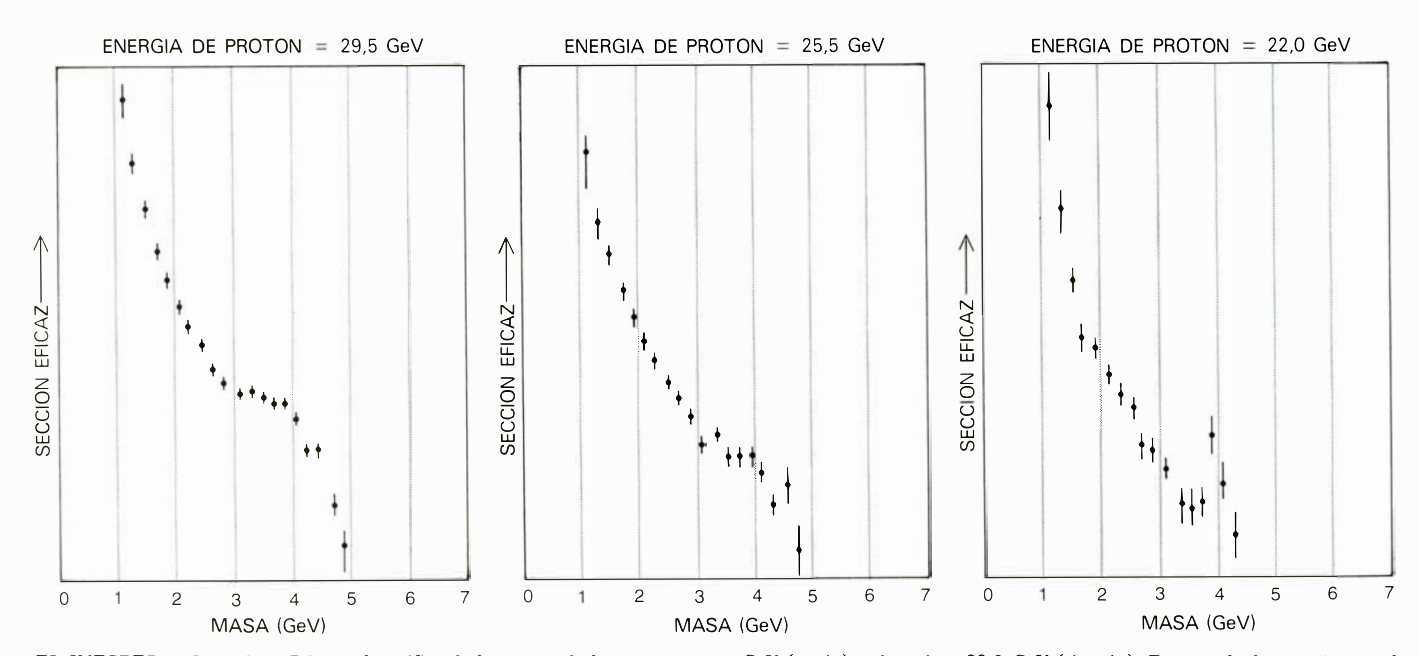


ESTE APARATO, en el acelerador de 30 GeV en el Brookhaven National Laboratory, generaba pares muónicos cuando los protones chocaban con uranio. Los muones pasaban a través del hierro, que absorbía los residuos nucleares no deseados. Los hodoscopios medían los ángulos y el momento de los muones.

casos pares leptónicos del abundante fondo de hadrones.

Tras darle muchas vueltas, nos dimos cuenta de que podríamos construir un sistema de detección basado en el hecho fundamental de que los leptones pueden penetrar en la materia, propiedad de la cual carecen los hadrones. Puesto que los muones pueden penetrar más en la materia que los electrones, decidimos concentrarnos en ellos e ignorar los pares electrón-positrón que también se producirían. Ello nos condujo a colocar tres metros o más de hierro entre el blanco de uranio y el detector de pares leptónicos. El hierro absorbería las partículas de interacción fuerte, si bien permitiría pasar a los muones y disparar una serie de contadores de centelleo.

Este sistema de detección presentaba el inconveniente de que podría alterar las trayectorias de los pares muónicos. Los átomos del hierro no sólo frenarían los muones, causándoles pérdida de energía, sino que repelerían y atraerían su carga eléctrica, desviándolos de sus caminos originales. Nos encontrábamos en un apuro. Si medíamos la energía y el ángulo de separación de los muones después de salir del absorbente de hierro y usábamos estos valores para calcular la masa de la partícula virtual que los producía, obtendríamos una contestación inexacta. Por otra parte, no cabía un cálculo más seguro observando los muones antes de penetrar en el absorbente, ya que el enorme flujo de hadrones impediría el cómputo de los leptones. Pero en esta primera fase de nuestro trabajo no nos preocupaba demasiado tener que conformarnos con una estimación imprecisa. Nuestro objetivo era detectar resonancias de masas altas no des-



EL INESPERADO "HOMBRO" en la gráfica de las masas de los protones virtuales generados en Brookhaven; tal protuberancia no desaparecía cuando la energía de los protones incidentes disminuía de 29,5 GeV (*izquierda*) a 25,5 GeV (*medio*) y después a 22,0 GeV (*derecha*). Este resultado sugería que el hombro era real, quizá se trataba de la resonancia, pobremente resuelta, de una nueva partícula y no el producto espúreo de un mal funcionamiento del aparato.

cubiertas aún, y creíamos que nuestro aparato era capaz de hacerlo aun cuando distorsionaría las características de tales resonancias. Valía la pena pagar ese pequeño tributo con tal de descubrir una nueva partícula.

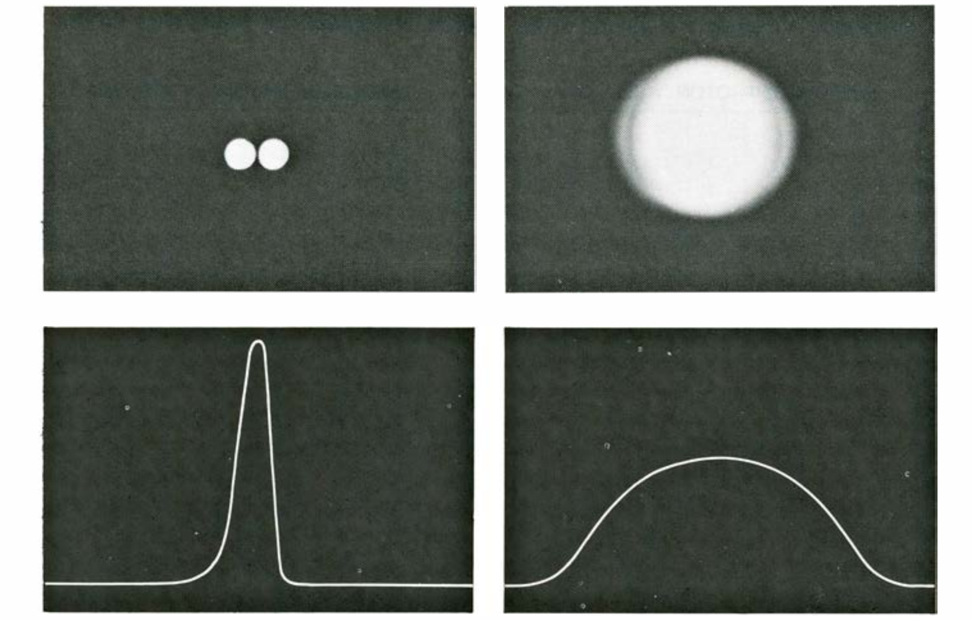
Empezamos la recogida de datos en el otoño de 1968. Un ordenador digital procesaba la información y dibujaba una gráfica de la cantidad de pares muónicos que se observaban para cada masa. Puesto que estábamos estudiando una reacción no explorada antes, no teníamos ninguna idea de cómo serían las distribuciones. Sin embargo, quedamos asombrados por la caída que empezaba a unos 1,5 GeV, se aplanaba justo por encima de los 3 GeV y se hundía luego muy rápidamente en el límite superior de nuestro sistema de detección, donde no podíamos recoger muchos datos [*véase la ilustración inferior de la página anterior*]. Este "hombro" de la curva nos excitaba. Nos preguntábamos si podría representar una resonancia estrecha que era difuminada por nuestro tosco aparato, pero que marcaba la presencia de alguna nueva partícula. Al disminuir la energía de los protones de tal bombardeo, la protuberancia no desaparecía. Eso constituía una buena señal. Significaba que dicha curiosa distribución no era, a buen seguro, resultado falso de algún capricho no detectado en nuestro equipo. Seguíamos afanados en la búsqueda de una prueba. No podíamos eliminar totalmente la posibilidad de que los efectos de distorsión del aparato fueran tan importantes que hubieran deformado la distribución de bajas energías, llegando a falsearla. Más aún, hubimos de considerar la posibilidad de que la protuberancia fuera una característica peculiar de una distribución suave de fotones virtuales en lugar de la resonancia difuminada de una nueva partícula.

A pesar del carácter provisional de nuestros resultados no interpretados, los teóricos los consideraron inmediatamente de interés porque parecían estar relacionados con la hipótesis de los quarks. La hipótesis original de 1964 sugería que todos los hadrones conocidos constaban de tres quarks, simbolizados por u, d y s [iniciales de "up" (arriba), "down" (abajo), "strange" (extraño)] y los tres antiquarks correspondientes $\bar{u}$, $\bar{d}$ y $\bar{s}$. Aunque el modelo de quarks original explicaba de forma elegante y simple las propiedades estáticas de más de 100 hadrones, no describía sus propiedades dinámicas. Sin embargo, en 1968, algunos pioneros habían usado el modelo de quarks para explicar datos de dispersión entre partículas y procesos de colisión. La principal dificultad de su explicación era su falta de unicidad: otras hipótesis razonables y distintas que no incorporaban los quarks podían explicar igualmente las características dinámicas.

Nuestros datos sobre pares leptónicos contribuyeron de forma notable a corroborar la explicación mediante quarks de la dinámica de los hadrones. En 1970, dos físicos de la Universidad de Stanford, Sidney D. Drell y Tung-Mow Yan, intentaron usar un modelo quark para generar de forma teórica nuestros resultados sobre pares leptónicos. Sus predicciones eran coherentes con nuestros datos alrededor de los 2 GeV, pero eran menores que éstos cerca de los 3 GeV. Animados por esta correlación parcial, por la intrigante posibilidad de una protuberancia y por el tremendo interés de los teóricos en nuestros datos, decidimos llevar a cabo una versión mejorada de nuestra experiencia en el acelerador, más potente, del Fermilab. La enorme energía de este acelerador, que en aquel entonces era de 300 GeV, aumentaría la probabilidad de emisión de pares a 3 GeV y esperábamos que esto nos permitiría, al fin, identificar el significado del misterioso hombro de la gráfica.

Luego, en 1974, y antes de que nosotros empezáramos a tomar datos, el modelo de tres quarks fue derrocado por lo que se llamó "la revolución de noviembre". El descubrimiento de una nueva partícula fue anunciado independientemente por Samuel C. C. Ting, de Brookhaven y el Instituto de Tecnología de Massachusetts, y por Burton D. Richter, de Stanford. En Brookhaven la nueva partícula, a la que Ting llamó J y Richter ψ (psi), aparecía como un pico espectacular en las masas de los fotones virtuales que se habían desintegrado en pares electrón-positrón [*véase la ilustración de la página 36*].

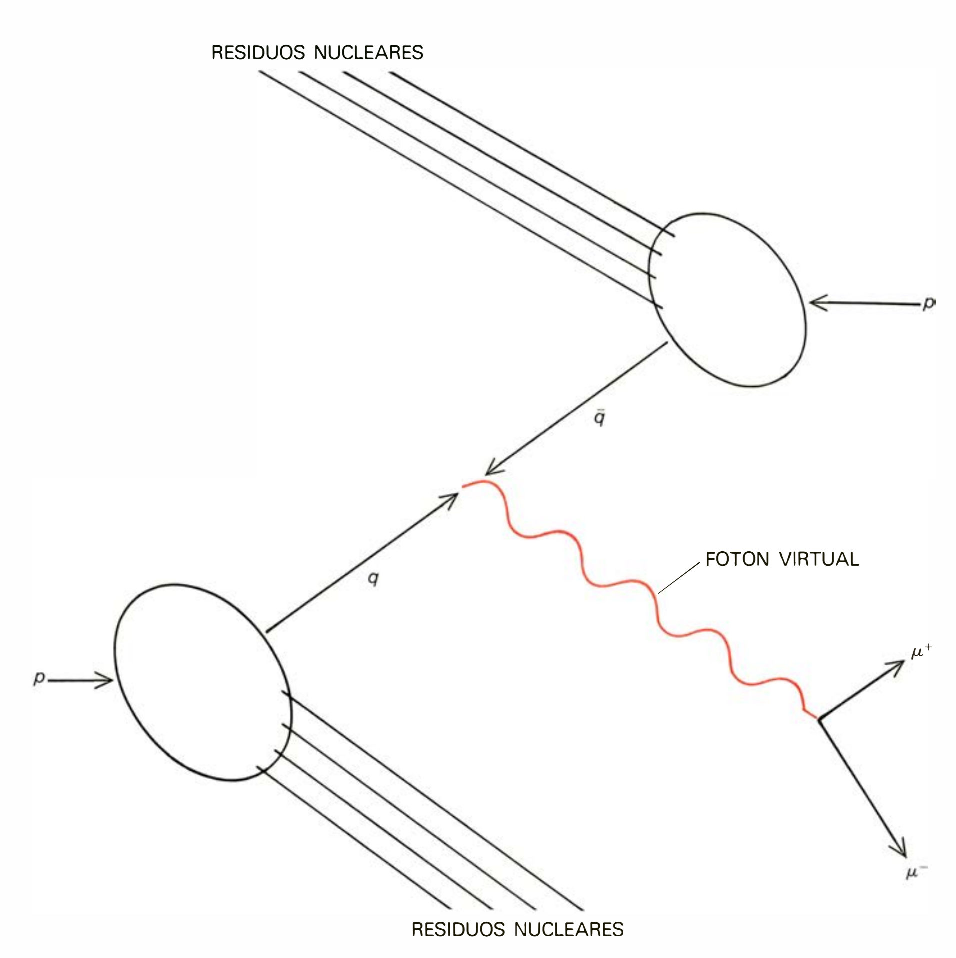
El descubrimiento de la J/ψ resolvía varios problemas importantes en la física de partículas. Explicaba nuestros datos sobre pares leptónicos y sugería la existencia de un cuarto quark, que se representa por c (inicial de "charm", palabra inglesa que significa encanto, y que es el nombre dado a la nueva propiedad mecánico cuántica que implicaba). La protuberancia que nosotros habíamos visto en el año 1968 se interpretaba ahora como una versión mal difuminada del estrecho pico a 3,1 GeV correspondiente a la J/ψ. Los aspectos revolucionarios de la J/ψ estribaban en que el pico era muy estrecho. De acuerdo con el principio de indeterminación de Heisenberg, una masa bien definida implica una vida media que es larga comparada con muchas de las otras partículas subatómicas. Y una vida media tan larga significa que la desintegración de la J/ψ en otras partículas (piones y kaones) estaba inhibida. La existencia de un cuarto quark explicaría este hecho. Puesto que los quarks son realmente fundamentales, una clase de quark no puede cambiar con facilidad en otra. Si las J/ψ estuvieran constituidas sólo por quarks encantados, no podrían desintegrarse fácilmente en piones y kaones, formados por los otros tres tipos de quarks. Las investigaciones que se realizaron a continuación confirmaron la interpretación de la J/ψ como un estado ligado de un cuarto quark y de su antiquark. El concepto de un cuarto quark, el encantado, adquirió aún más peso cuando se



EL PROBLEMA DE UNA RESOLUCION POBRE en los experimentos de alta energía queda ilustrado comparando las fotografías con las curvas de resultados experimentales. Una cámara desenfocada puede hacer que dos luces *(arriba, a la izquierda)* **aparezcan como una sola** *(arriba, a la derecha)*. **El aparato del Fermilab hacía que la resonancia estrecha de la partícula J/ψ** *(abajo, a la izquierda)* **apareciera como ancha** *(abajo, a la derecha)*. **Un ordenador ayudaba a clarificar la distorsión. (La J/ψ se descubrió en 1974.)**

descubrieron partículas que parecían estar formadas por varias combinaciones de los cuatro quarks.

Al comparar el hombro de nuestra gráfica de 1968 con los datos de 1974 referentes a la J/ψ observamos que lo que habíamos ganado en sensibilidad, lo habíamos perdido en resolución. Habíamos detectado más de 10.000 pares muónicos con nuestro aparato, que era muy sensible, pero no pudimos interpretar la distribución difuminada. Los descubridores de la J/ψ en Brookhaven, por otra parte, usando una nueva generación de detectores de partículas detectaron sólo 242 pares, pero como su aparato podía localizar la posición de los pares en la escala de masas con mayor exactitud vieron un pico estrecho y muy destacado.



MODELO QUARK, introducido para explicar las propiedades estáticas de las partículas; puede dar cuenta también de ciertos procesos dinámicos, verbigracia, la creación de pares muónicos. Sidney D. Drell y Tung-Mow Yan, de la Universidad de Stanford, propusieron que un fotón virtual que se desintegra en un par muónico se forma cuando un quark (q) del protón incidente y un antiquark ($\bar{q}$) del "mar" de quarks asociado al nucleón que actúa como blanco se aniquilan entre sí. Drell y Yan intentaron predecir los datos de Brookhaven y tuvieron éxito con masas próximas a los 2 GeV, pero no con otras cercanas a los 3 GeV.

Ahora que quedaba resuelto el misterio del hombro, decidimos usar el nuevo acelerador del Fermilab para buscar resonancias en la región no explorada de masas por encima de los 5 GeV. En 1975 y 1976, en tres tomas de datos, observamos cientos de pares leptónicos. La energía del acelerador del Fermilab se había aumentado hasta 400 GeV, un incremento que iba a ser crucial para nuestro trabajo. Esta vez íbamos a poder calibrar los efectos distorsionadores de los aparatos examinando cómo se alteraba la resonancia J/ψ, lo que no hubiéramos podido hacer en 1968. Teníamos años de experiencia en el estudio de pares muónicos y de progreso en el desarrollo de detectores del que podríamos hacer buen uso.

En febrero del año pasado, nuestro grupo empezó a montar la nueva versión del experimento de pares leptónicos utilizando lo que habíamos aprendido en los dos años anteriores. Nos dimos cuenta de que, para sacar conclusiones acerca de masas mayores y más raras, debíamos observar muchos más sucesos. Al mismo tiempo, había que mejorar la resolución o nos encontraríamos con los mismos datos no interpretables que habíamos recogido en 1968.

John Yoh, de Columbia, había observado un pequeño número de sucesos en los entornos de 9,5 GeV en nuestros resultados de 1976. Puso una botella de champagne Moët, etiquetada "9,5", en la nevera de nuestro grupo. Eso no convenció a nadie de que estuviéramos en la pista de una nueva partícula. Pero nos alentaba en nuestra búsqueda el que nuestros datos eran únicos: nadie más había nunca visto 350 pares leptónicos con masas mayores de 5 GeV. Quizás había algo allí, en algún sitio.

La experiencia nos demostró que podíamos acercar más los detectores al blanco, de suerte que detectaran más pares muónicos. Stephen W. Herb, de Columbia, había predicho acertadamente que esto no aumentaría también el enorme flujo de hadrones que interferían con el sistema de detección. En 1968 habíamos usado hierro para absorber las partículas no deseadas, pero los átomos de hierro con sus 26 protones y 26 electrones ejercían una fuerza electromagnética sobre los muones que los desviaba de sus caminos originales. El resultado era que las masas de los pares muónicos no podían ser calculadas con precisión. Ahora usaríamos berilio metálico como absorbente principal. Con sólo cuatro protones y cuatro electrones, el berilio casi no desviaría los pares muónicos, aunque fuera capaz de eliminar la mayoría de los hadrones.

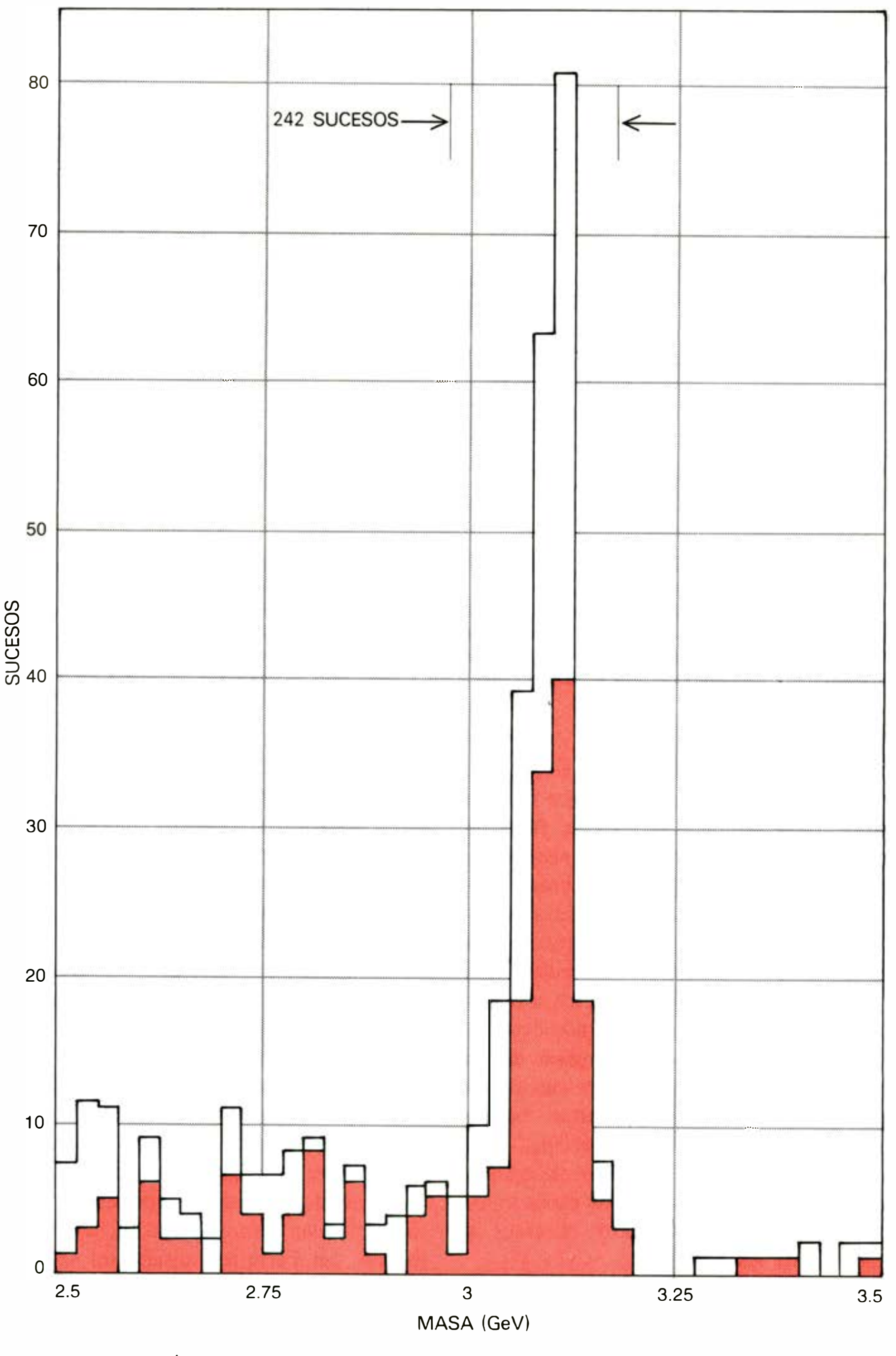
Quedaba aún por resolver una dificultad importante. Los muones son unas partículas traicioneras. Como ya he mencionado, pueden penetrar fácilmente a lo largo de muchos metros de hierro. Teníamos que estar totalmente seguros que nuestros pares muónicos eran tales, es decir, que habían sido creados en el blanco y que desde allí se habían dirigido sin desviarse y sin chocar a través del gran imán deflector hasta nuestros contadores. Para ganar seguridad queríamos colocar un detector en el centro del imán. La cosa resultaba difícil (era como diseñar un reloj delicado y preciso para que funcionara en un horno de fundición), pero Walter R. Innes, del Fermilab, logró un esquema adecuado. No nos acababa de satisfacer la instalación. Los sucesos más interesantes eran también los menos frecuentes. Después de una larga jornada con más de 10.000 millones de colisiones nucleares por segundo, sucesos extraordinariamente improbables podrían conspirar para dar un resultado falso de la experiencia. Para eliminar esta eventualidad, Charles N. Brown, del Fermilab, diseñó un sistema magnético sencillo que volvería a medir la energía de cada muon a su salida del detector principal.

El primero de mayo del año pasado obtuvimos nuestros primeros datos. Nos llenó de entusiasmo descubrir que el aparato mejorado registraba 90 veces más pares muónicos que un año antes. El acelerador con la energía aumentada fun-

cionaba soberbiamente, suministrando cantidades ilimitadas de protones con gran precisión. En la primera semana observamos 3000 pares muónicos cuyas energías superaban los 5 GeV, más de 10 veces los datos del resto del mundo, y de mucho mejor calidad. Representamos los resultados y parecían notablemente libres de los efectos de interferencia de los hadrones. La resonancia J/ψ aparecía con nitidez, lo cual significaba que habíamos resuelto con éxito el problema de aumentar el poder de resolución de nuestro aparato. Nuestra excitación alcanzó un máximo cuando vimos que el descenso uniforme en la producción de pares muónicos al aumentar su masa quedaba interrumpido cerca de 10 GeV por una intrigante protuberancia.

La semana siguiente doblamos el número de datos y la protuberancia continuaba en su lugar. Aunque ya no la podíamos descartar como un suceso casual engañoso, nos preguntábamos si podía tratarse del producto caprichoso de alguna peculiaridad no detectada de nuestro aparato. Afortunadamente, disponíamos de mecanismos para ver si esto era así. Miramos, uno tras otro, cada centímetro cuadrado de la superficie del detector para ver cómo estaban distribuidos los muones que habían llegado a cada área. Por todas partes encontramos distribuciones suaves, lo que indicaba que el aparato no había generado la resonancia. Aún más, cuando de forma artificial mezclamos los μ^+ del lunes con los μ^- del martes para formar falsos pares $\mu^+\mu^-$, obtuvimos distribuciones perfectamente suaves que estaban de acuerdo con todas nuestras expectativas de cómo debía funcionar el equipo. Después de hacer pasar el aparato por nuevas pruebas, y acumular más datos, llegamos al convencimiento de que la resonancia representaba algo real: una nueva partícula cuya masa rondaba los 10 GeV. Si bien queríamos mantener los resultados secretos hasta poderlos interpretar totalmente, los rumores de nuestro descubrimiento se propagaron en seguida entre la comunidad de físicos. Por tanto, el día 20 de junio dimos publicidad a nuestros datos: 26.000 pares, casi 100 veces el número de datos combinados de todas las experiencias anteriores. Llamamos a esta partícula hypsilón.

A continuación nos propusimos determinar la anchura de la resonancia, usando el mismo método por el que se había medido la de la J/ψ. La anchura es la indeterminación en la masa de la resonancia; el principio de indeterminación de Heisenberg asocia un pico estrecho (una indeterminación pequeña) con una vida media larga y un pico ancho (una indeterminación grande) con una vida media corta. Después de reunir más datos encontramos que la resonancia constaba de dos picos próximos (y quizás un tercero) separados por 600 MeV (600 millones de electronvolts) y cada pico tenía 500 MeV de anchura. Esto indicaba que la hypsilón podía existir en dos o quizá tres estados de energías ligeramente distintas. Antes de que pudiéramos concluir, a partir de estas anchuras, que la resonancia hypsilón era intrínsecamente estrecha debíamos tener en cuenta los efectos de distorsión del aparato inevitablemente imperfecto. Los aparatos con un poder de resolución bajo hacen que los picos de la naturaleza aparezcan anchos, de forma análoga a como un sistema óptico de pobre calidad hará borrosos los detalles finos de una fotografía. En nuestro experimento de Brookhaven de 1968, una pobre resolución había distorsionado los datos hasta el punto de hacerlos no interpretables.



LA PARTICULA J/ψ se descubrió en 1974 como una estrecha resonancia en las masas de los fotones virtuales que se desintegraban en pares electrón-positrón. La distribución coloreada representa la distribución de masas que se obtenía cuando el espectrómetro de detección funcionaba con partículas incidentes a la intensidad normal; la distribución en blanco correspondía al caso en que se rebajaba la intensidad en un 10 por ciento. La resonancia a 3,1 GeV, que aparecía claramente en ambas circunstancias, se interpretó como una versión, con un poder de resolución mucho más elevado, de la protuberancia hallada en Brookhaven en 1968. La partícula J/ψ indicaba la existencia de un cuarto quark, representado por c, correspondiente a "charm" (encanto), y el correspondiente antiquark, que se representa por $\bar{c}$. (Ilustración de Allen Beechel.)

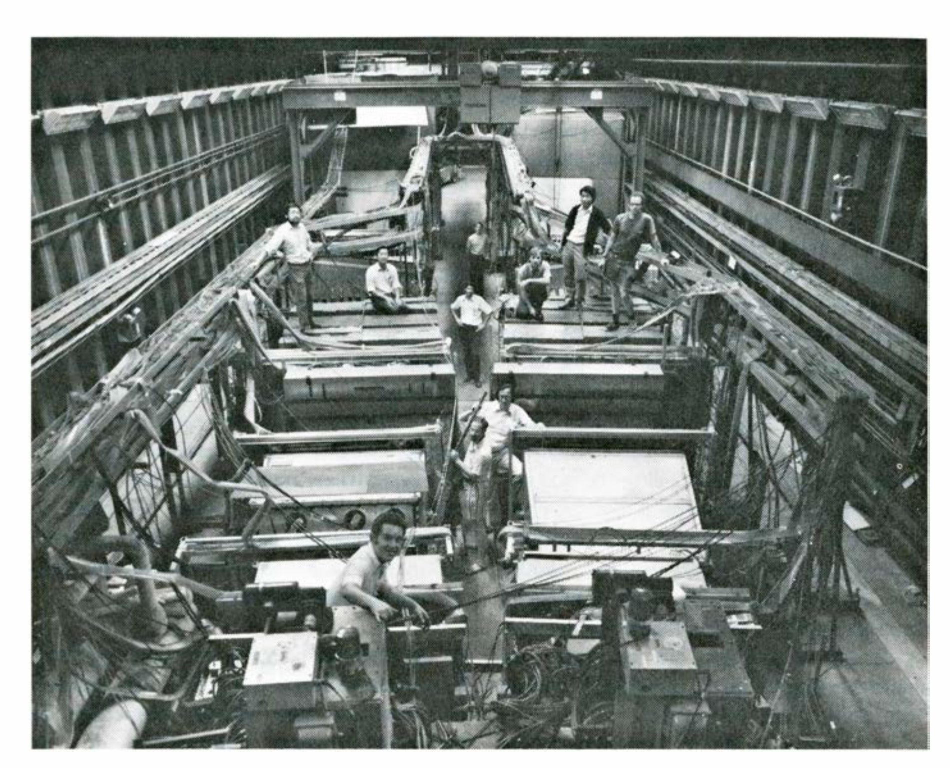
Para determinar cómo el aparato del Fermilab había deformado la resonancia, utilizamos el método de simular con un ordenador todo nuestro experimento. Tales simulaciones, que se llaman métodos de Monte Carlo, se usan de manera habitual en física de altas energías. En nuestro caso, el ordenador, programado de forma que conociera la localización y función de cada pieza de nuestro aparato, seleccionaba una configuración de dos muones y trazaba sus trayectorias hasta el detector final. Si el programa del ordenador hacía que el electrón encontrara un absorbente, por ejemplo berilio, el programa simularía la forma como la producía en la realidad la colisión muon-berilio. Con un potente ordenador, podíamos simular la traza de decenas de miles de pares muónicos a través del aparato.

Representamos gráficamente la distribución de masa de los sucesos de Monte Carlo y descubrimos que la resonancia hypsilón así simulada era mucho más estrecha que la medida. Esto sugería que la anchura medida se producía, sobre todo, por el aparato. Los programas complicados de ordenador pueden tener fallos. ¿Los habíamos corregido todos? Quizá la simulación fuera errónea y la resonancia poseyera, en realidad, la anchura que habíamos medido. Por fortuna, teníamos medios de eliminar esta posibilidad. Puesto que ya sabíamos cómo nuestro aparato había distorsionado configuraciones tales como la J/ψ, podíamos comprobar nuestro programa de Monte Carlo para ver si revelaba correctamente tales distorsiones. Así fue; y concluimos, con plena seguridad, que la resonancia hypsilón tenía una anchura menor de 100 MeV. Esta anchura extraordinariamente pequeña indicaba que la nueva partícula gozaba de una vida media muy larga.

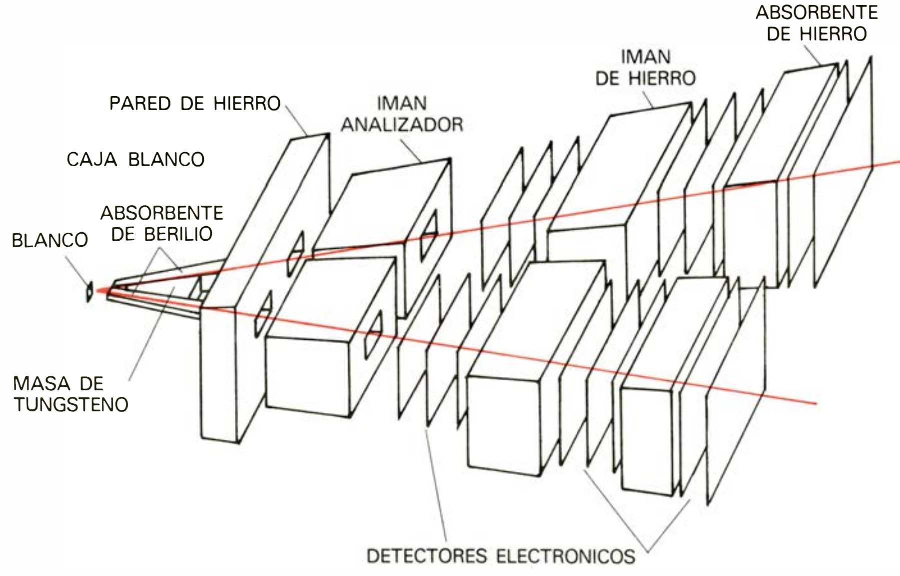
En situaciones normales sería de esperar que una partícula cuya masa fuera nada menos que 10 veces la masa del protón tuviera una enorme cantidad de estados de masa inferior en los que pudiera desintegrarse, contribuyendo cada estado a disminuir su vida media. Pero contrariamente a lo esperado, la hypsilón, la partícula más pesada que se ha descubierto, tiene una vida media larga. Esto significa que no se desintegra en los hadrones de masa menor, todos los cuales están formados por los quarks *u, d, s* y *c*. Sin embargo, cuando el año pasado tuvo lugar nuestro descubrimiento no se conocía ninguna ley de la física que explicara este hecho. La conclusión era clara y excitante: alguna nueva ley de física prohíbe (o con más precisión, inhibe) que la hypsilón decaiga en hadrones ordinarios.

En busca de esta nueva ley miramos si en algún trabajo teórico se había previsto el descubrimiento de la hypsilón. A lo largo de los años los teóricos habían sugerido la existencia de nuevas partículas para dar cuenta de datos enigmáticos, y nos preguntábamos si la hypsilón pudiera ser una de tales partículas.

El único candidato razonable era un nuevo quark, más pesado, ligado a su antiquark en configuraciones similares a átomos que aparecían como un conjunto de masas muy cercanas unas a otras y con características parecidas a la hypsilón. Los trabajos teóricos que habían sugerido la posibilidad de un tal quark eran divagaciones especulativas basadas en prejuicios estéticos. En un grupo de trabajos se esperaba que la existencia de un nuevo quark sería capaz de explicar algunos resultados curiosos de ciertas experiencias de colisio-



APARATO DEL FERMILAB, que detectó la partícula hypsilón. Los protones incidentes y los nucleones blanco chocan en un punto, justo fuera de la figura, del primer plano. Los electroimanes, localizados a la izquierda y a la derecha enfrente del hombre del primer plano, desvían los pares muónicos de suerte que puedan medirse sus energías y ángulos de separación. Las dos componentes de cada par muónico viajaban a través de distintos brazos de los detectores. Los dos brazos, de un 1,8 metros de anchura, 1,8 altura y 30 metros de longitud, se extendían desde los electroimanes hasta donde está el hombre del fondo de la imagen.



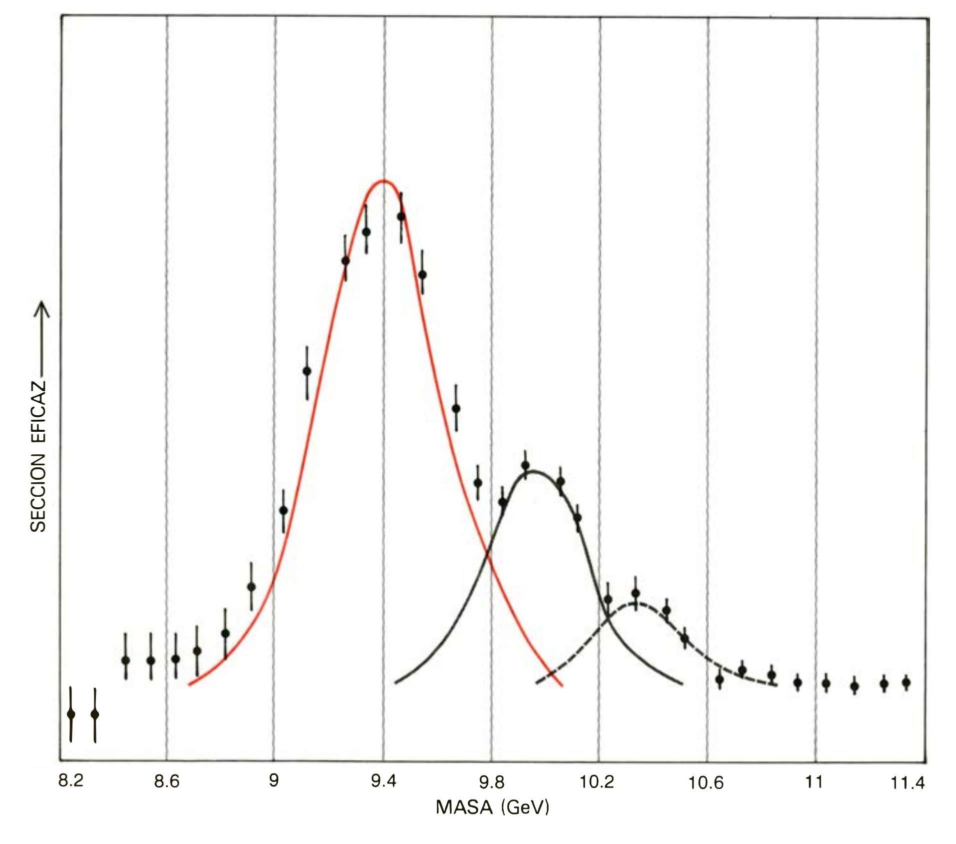
SE USO UN ABSORBENTE DE BERILIO en el experimento del Fermilab para apantallar los productos nucleares no deseados, ya que afecta a las trayectorias de los muones mucho menos que el hierro. La masa de tungsteno servía para recoger los protones incidentes que no habían chocado con el blanco de nucleones. El esquema constituye una vista lateral del aparato, contemplado desde arriba, de la fotografía superior.

nes de neutrinos. Nuestros mejores cálculos, realizados al principio del presente año, indicaban que la hypsilón tiene resonancias a 9,4, 10,0 y 10,4 GeV [*véase la ilustración inferior de esta misma página*]. Una partícula constituida por un quinto quark y su antiquark podría existir en el estado fundamental, o estado más bajo, a 9,4 GeV, y en estados excitados a 10,0 y 10,4 GeV. Además, la existencia de una quinta subpartícula explicaría de una forma elegante la larga vida media de la hypsilón, de forma análoga a como el cuarto quark había explicado la larga vida de la J/ψ. Si la hypsilón estaba formada únicamente por la quinta clase de quarks, no podría desintegrarse en hadrones ordinarios que constan de distintas combinaciones de los otros cuatro quarks. Todas estas consideraciones de peso convencieron a la mayoría de los físicos de partículas de que la hypsilón era, en efecto, un objeto compuesto de un quark de la quinta clase unido a su antiquark.

En la historia de la hipótesis de un quinto quark quizá muchos no sepan que las razones por las que algunos teóricos lo introdujeron resultaron, más tarde, superfluas. Se suponía que la nueva subpartícula explicaba algunos datos enigmáticos de experiencias de colisiones, los cuales, examinados con más detalle, no resultaron enigmáticos en absoluto. Donde el cuarto quark había explicado toda clase de fenómenos enigmáticos, el quinto quark explicaba únicamente los resultados de nuestro experimento. Los datos de colisiones mal interpretados habían servido, sin embargo, para estimular la especulación en torno a las propiedades de las partículas compuestas de quarks más pesados, propiedades que tenía la hypsilón. El hecho de que los cuatro quarks identificados se apareasen como "up" y "down" y como "strange" y "charm", condujo a los teóricos a predecir que, si había un quinto quark, debería existir también un sexto. Los excéntricos nombres de "top" (pico) y "bottom" (fondo) o "truth" (verdad) y "beauty" (belleza) habían quedado reservados para los dos nuevos quarks, en caso de que se descubrieran.

Las resonancias de la partícula hypsilón presentaban a los físicos una riqueza perturbadora: una familia inesperada de partículas nuevas compuestas de un quinto quark imprevisto. La hypsilón había causado una gran conmoción. Desencadenó búsquedas de otras partículas pesadas en zonas inexploradas de masas, y arrojó luz sobre la inescrutable fuerza fuerte. Esta fuerza, que liga los quarks para formar hadrones y los hadrones para formar núcleos atómicos, es demasiado intensa para poderla investigar mediante las técnicas convencionales de colisiones. Sin embargo, cualquier modelo de fuerza fuerte que se proponga, al ser una descripción de la fuerza entre un quark y un antiquark, debe predecir correctamente los niveles energéticos (o masas) de la familia hypsilón. En la hypsilón, a diferencia de lo que sucede con hadrones de masa menor, es fácil calcular estas predicciones, ya que las velocidades de los quarks con gran masa son relativamente bajas. Lo que significa que las complicadas consideraciones relativistas no entran en el cálculo. Las predicciones de varios de tales modelos teóricos han fallado, por lo que nos eximen de su consideración. Los modelos que han tenido éxito sugieren todos que el nuevo quark presenta una carga $-1/3$ (siendo la carga del electrón -1) y que la fuerza entre quarks aumenta al crecer la distancia entre ellos. Se ha propuesto esa fuerza para justificar el fracaso de los físicos de partículas a la hora de observar quarks en estado libre: se requiere tal cantidad de energía para lograr su separación que, cuando ésta se suministra, se gasta en crear nuevos pares de quarks, en vez de aprovecharse para separar los antiguos. La idea generalizada hoy es que los quarks están permanentemente confinados en estructuras compuestas.

La confirmación de esta hipótesis y una mejor comprensión de las fuerzas fuertes pueden, sin embargo, no ser consecuencia de las partículas hypsilón creadas en los aceleradores de protones, tales como los que nosotros hemos utilizado en Brookhaven y en el Fermilab. El proceso de producción resulta demasiado complicado. Por ejemplo, en la colisión inicial en el Fermilab intervienen tres quarks del proyectil protón que chocan con tres quarks del nucleón blanco. El próximo paso para desvelar el misterio de las fuerzas entre quarks se realizará cuando los físicos sean capaces de estudiar intensivamente las partículas hypsilón producidas en máquinas de anillos de almacenamiento en las que electrones y positrones circulan en direcciones opuestas y chocan los unos con los otros. Estas máquinas nos darán datos más claros y con mayor poder de resolución. En realidad, ya lo han hecho. En abril, un grupo de físicos del DESY (siglas de Deutsches Elektronen Synchroton), en Hamburgo, modificó su anillo de almacenamiento electrón-positrón para buscar la partícula hypsilón. La encontraron a 9,46 GeV y pudieron poner un límite superior de 7 MeV a la anchura de esta resonancia. (Esto es una mejora notable sobre nuestro valor de 100 MeV.) Sus datos también sugerían una carga $-1/3$. Es de presumir que la espectroscopía del



TRES RESONANCIAS MUY PROXIMAS caracterizan la partícula hypsilón. Con un estado más bajo de 9,4 GeV y estados de 10,0 y 10,4 GeV, la partícula se interpretó como formada por un quinto quark, de gran masa, ligado a su antiquark. La espectrocopía de la hypsilón será clasificada mediante experiencias llevadas a cabo en aceleradores de anillos de almacenamiento. La búsqueda de un sexto quark ha empezado ya.

PARTICULA	SIMBOLO	MASA (GeV)
PROTON	γ	0
NEUTRINO	ν	0
ELECTRON	e	.0005
MUON	μ	.105
MESONES PI	π^0	.135
	$\pi^\pm$	.140
MESONES K	$K^\pm$	.494
PROTON	p	.938
NEUTRON	n	.940
FI	ϕ	1.020
LAMBDA	Λ	1.116
MESONES ENCANTADOS	D^0	1.863
	D^+	1.868
LAMBDA ENCANTADO	Λ_c	2.260
FAMILIA J/ψ	J/ψ	3.098
	ψ'	3.684
FAMILIA HYPSILON	Υ	9.4
	Υ'	10.0
	Υ''	10.4

MASAS DE LAS PARTICULAS dadas en orden creciente. La hypsilón con una masa en el estado fundamental que equivale a una energía de 9,4 GeV, es la más pesada de las partículas hasta ahora descubiertas. Pesa tres veces más que la J/ψ, que es la segunda partícula de mayor masa, y 19.000 veces más que el electrón. Las partículas que incorporaran el sexto quark serían aún más pesadas que la hypsilón.

sistema hypsilón se obtenga con detalle en los próximos años, cuando los nuevos anillos de almacenamiento de mayor energía, que se están construyendo en Hamburgo, Stanford y en la Universidad de Cornell, hayan sido terminados. Los físicos buscarán el sexto quark, la esperada pareja del quinto, y entonces intentarán descubrir partículas formadas por combinación de los seis quarks.

Al ir avanzando las técnicas de aceleración de partículas los físicos continuarán descubriendo, sin lugar a dudas, nuevas entidades subatómicas. Su proliferación originará profundas e inquietantes preguntas: ¿Están limitadas, en cuanto a número, las clases de quarks? ¿Si hay seis, por qué no 12? ¿Si hay 12, por qué no 24? ¿Y si el número de quarks es elevado tiene algún sentido llamar a los quarks elementales? La historia de la ciencia sugiere que la proliferación de entidades físicas constituye una prueba tangible de que estas entidades no son elementales. Los químicos del siglo XIX redujeron la variedad, aparentemente infinita, de substancias químicas a unos 36 elementos, los cuales en el transcurso de los años han sobrepasado la cifra de 100. Como últimos e indivisibles constituyentes de la materia, los elementos químicos resultaron ser demasiados. En los 1930 se descubrió que todos los elementos estaban formados por electrones, protones y neutrones. Después de la Segunda Guerra Mundial, se unieron a estas partículas docenas de otras: piones, kaones, lambdas, etcétera. De nuevo, había demasiadas. Entonces pareció que todas éstas reducían a tres quarks. Ahora la experiencia demuestra que existe un cuarto y un quinto quark. ¿Son también demasiados? ¿Se propondrán pronto estructuras más simples como constituyentes de los quarks? ¿Es posible que no haya partículas realmente elementales, que cada entidad de la naturaleza tenga sus partes constituyentes? ¿O estará la simplicidad final en la que creen muchos físicos alojada en grupos matemáticos que ordenan las partículas más que en objetos genuinamente elementales?

Dejando de lado estas últimas ideas de refinada elucubración, muchos físicos desconfían de tales especulaciones ante la dificultad, si no imposibilidad, de examinar un quark aislado. Sin embargo, la experiencia con la partícula hypsilón indica que, a pesar de esta dificultad, puede obtenerse un conocimiento detallado de los movimientos y fuerzas entre los quarks. La aparente inseparabilidad de estas entidades no debe entorpecer el camino de la investigación. Fijémonos en la lección que los físicos sacaron del electrón. El desarrollo de la teoría del electrón hubiera sido probablemente más lento, aunque no inviable, si los electrones hubieran sido observados únicamente ligados a los átomos y nunca en estado libre. Esta es la respuesta de un experimental a los profetas del catastrofismo que ven el confinamiento de los quarks como un último límite del conocimiento: una valla levantada por la naturaleza para ocultar sus últimos secretos por siempre jamás. ¿Y quién puede afirmar que los físicos nunca construirán un acelerador superpotente, capaz de vencer la fuerza de confinamiento y liberar el quark?

Ciencia y sociedad

Estimulación precoz en subnormales

El síndrome de Down, conocido vulgarmente con el nombre de mongolismo, es el trastorno cromosómico más frecuente asociado con el retardo en el desarrollo, ya que se estimaque afecta a un 10 por ciento de los niños diagnosticados como subnormales. Dicho trastorno suele consistir en la existencia de una trisomía –es decir, el exceso de un cromosoma– en el par 21; mucho más raros son los casos de traslocación, en los que el cromosoma suplementario está adherido a otro cromosoma, o los de mosaico, en los que las células con trisomía 21 se alternan con células normales.

Se trata de la primera alteración cromosómica descubierta, y fue puesta de manifiesto por Lejeune, Gautier y Turpin en 1959. Aunque se admite, corrientemente, que aparece un niño con síndrome de Down por cada 650 nacidos vivos, en realidad, dado que este trastorno cromosómico se encuentra estrechamente relacionado con la edad a la que las madres dan a luz –la probabilidad pasa de 1: 2500 para madres de menos de 30 años a 1: 32 para madres de más de 45 años– y teniendo en cuenta que la media de dicha edad varía según los países y las épocas, tal estimación sólo es válida para determinadas culturas y momentos históricos. En la actualidad, por ejemplo, con las prácticas reproductoras vigentes en las sociedades respectivas, el riesgo de tener en España un niño mongólico es doble que en Suecia y triple que en el Japón.

Hasta fechas muy recientes, las víctimas de esta anomalía genética tenían, desde el punto de vista intelectual, un pronóstico marcadamente pesimista, aconsejándose a los padres, en bastantes casos, su institucionalización. Diversos autores señalan que la edad mental de las personas afectadas correspondientes a grupos de todo tipo de edad oscila entre 1 y 8 años. En 1969, Benda avanzó que el niño mongólico suele tardar 10 años en conseguir lo que los niños normales adquieren en tres.

Durante el verano de 1978 se celebró en Barcelona un seminario sobre estimulación precoz en niños con síndrome de Down, a cargo de Valentine Dmitriev, coordinadora del programa experimental que se lleva a cabo en uno de los centros de la Universidad de Washington. También participó activamente en la dirección del seminario Moira Pieterse, coordinadora de un programa similar que, desde 1975, se desarrolla en la Universidad Macquarie de Sidney. Los datos aportados, fruto de un intenso y cuidadoso trabajo que continúa en la actualidad, tiñen de moderado optimismo las perpectivas de futuro de los niños aquejados por esta anomalía genética.

El programa de la Universidad de Washington actualmente en curso empezó en enero de 1971 con 11 niños comprendidos entre los 18 meses y los tres años de edad; más tarde, durante el curso escolar 1973-74, se incorporaron al mismo otros 50 niños con síndrome de Down de edades comprendidas entre el nacimiento y los seis años y medio. Hacia el otoño de 1974, varios de estos niños habían ya completado su curriculum preescolar y se encontraban preparados para incorporarse a la escuela primaria.

Mientras que los niños con síndrome de Down, de edad comprendida entre el nacimiento y los 12 meses, que no han pasado por el programa experimental de la Universidad de Washington sólo son capaces, por término medio, de efectuar un 62 por ciento de las tareas que realizan los niños normales, los niños mongólicos que han sido sometidos a las técnicas de estimulación precoz de dicha Universidad, de edad comprendida entre los 60 y los 72 meses de edad, son capaces de llevar a cabo el 95 por ciento de las tareas que efectúan los niños normales.

Es importante destacar que esta pauta de desarrollo acelerado conseguida por los investigadores de la Universidad de Washington difiere de la obtenida por otros especialistas cuyos datos muestran un nivel de desarrollo relativo continuamente decreciente en los niños con síndrome de Down. En otras palabras, el programa de la Universidad de Washington muestra una alentadora relación positiva entre la edad física y el nivel de desarrollo; a medida que aumenta la edad física aumenta, paralelamente, el nivel de desarrollo.

Creemos que puede ser de interés señalar que las técnicas de estimulación precoz utilizadas se basan, fundamentalmente, en el paradigma de aprendizaje puesto en evidencia por el profesor de la Universidad de Harvard B. F. Skinner, y que los resultados obtenidos proporcionan un dramático apoyo fáctico a unas palabras de Throne pronunciadas hace ya algunos años. El descubrimiento de Skinner de que el comportamiento de los organismos –humanos o no– es función de sus consecuencias equivale, como indica Throne, a que "la conducta del retardado se encuentre limitada solamente por nuestro esfuerzo y habilidad para aplicar, de forma contingente, las infinitas consecuencias potencialmente disponibles para mejorarla". (R. B.)

Los mejillones y la contaminación

El mejillón, organismo micrógrafo, filtrador y sedentario, está adaptado a soportar condiciones muy adversas. Tolera estados de anoxia (falta de oxígeno) prolongados, durante la bajamar y a lo largo de su permanencia en los mercados, y muestra tenaz resistencia a factores físico-químicos que resultan mortales para la mayoría de las especies. Mientras algunos polutantes pueden estar presentes en el agua de mar en concentraciones del orden de 10^{-15} a 10^{-12} gramos de contaminante por gramo de agua, en los tejidos de *Mytilus edulis* se han registrado concentraciones del orden de 10^{-9} a 10^{-4} gg^{-1}. (Datos publicados por Goldberg y colaboradores en 1978.)

La gran capacidad filtradora del mejillón queda reflejada en las siguientes cifras: ejemplares de 70 a 80 milímetros de longitud son capaces de filtrar, en una hora, de medio litro a cuatro litros de agua de mar, capacidad filtradora superada por la de las ostras que en una hora llegan a filtrar hasta 24 litros. Este volumen de agua filtrada varía de acuerdo con una amplia gama de factores, como son la temperatura, la cantidad y calidad de elementos en suspensión (desde partículas inorgánicas a microorganismos planctónicos) y determinadas sustancias disueltas. Se ha comprobado que el mejillón prefiere, con mucho, la diatomea *Nitzschia* a la alga verde unicelular *Chlorella*, cuya presencia retarda la velocidad de filtración. Un estudio detallado sobre los distintos factores que influyen en la velocidad de filtración de los mejillones y otros

lamelibranquios puede hallarse en el trabajo de Phillips, publicado en *Environmental Pollution* (vol.13., pág. 281, 1977).

Los mejillones pueden retener partículas de 1 a 2 milimicras de diámetro, si bien no todo el material en suspensión de estas dimensiones es retenido por estos bivalvos; expulsa una parte del mismo a los pocos minutos en forma de cordones mucilaginosos, más o menos compactos, que reciben el nombre de pseudoheces.

Una experiencia muy ilustrativa de la formación de las pseudoheces y la retención de partículas la llevamos a cabo en el departamento de Morfología y Microscopía de la Facultad de Biología de la Universidad Central de Barcelona, como complemento de un estudio ultraestructural de gónadas femeninas de algunos moluscos (Durfort, 1973). Las células sexuales femeninas del mejillón tienen una membrana citoplasmática u oolema particularmente adaptada a favorecer al máximo los procesos de absorción, dado que está erizada de evaginaciones digitiformes o microvillis de una micra de longitud, así como de abundantes formaciones pinocíticas, puestas de manifiesto con el microscopio electrónico.

Nos interesaba comprobar la incorporación de partículas por parte de dichas células germinales, proceso implicado luego en la vitelogénesis de los ovocitos, para lo cual se procedió a colocar varios lotes de mejillones en tanques de plástico con agua de mar, convenientemente oxigenada, añadiendo a cada tanque dosis distintas de hierro coloidal (preparado en el laboratorio por diálisis de cloruro de hierro), o bien suministrándolo en forma de aspartato ferroso.

Los lotes sometidos a mayores concentraciones de hierro coloidal eliminaban, a los pocos minutos, considerables cantidades de pseudoheces azuladas, quedando el agua muy clara y transparente. Estudiando microscópicamente cortes de diversos órganos de los ejemplares de los distintos lotes, pudimos comprobar, tras reacciones citoquímicas muy elementales, que había habido incorporación de hierro por parte de las células del mejillón, siendo las branquias las primeras en dar reacción positiva, seguidas del hepatopáncreas y de los ovocitos. Los lotes sometidos a condiciones más extremas morían a los seis días de haber montado la experiencia.

Es interesante subrayar el hecho de que los parásitos que, en un tanto por ciento variable invaden el hepatopáncreas y las gónadas de los mejillones, son muy lábiles y fallecen a las pocas horas de encontrarse en presencia de concentraciones medianas de hierro coloidal.

El poder filtrador y clarificador de las aguas por parte de los moluscos bivalvos ha adquirido en los últimos años gran interés, dado que por desgracia son frecuentes los vertidos de aceites y petróleos en las aguas marinas. Recordemos el accidente ocurrido en mayo de 1976 al hundirse el Urquiola frente a las costas de La Coruña, más recientemente el buque libanés Amoco Cádiz frente a las costas bretonas y otro petrolero griego, Cristo Bitas, cerca de las costas galesas (octubre, 1978). Las ochenta mil toneladas de petróleo vertidas a 250 kilómetros de la costa de Bretaña, por el Amoco Cádiz, han contaminado cinco mil toneladas de ostras de unos tres años de edad, las de máxima cotización en el mercado, que se están destruyendo actualmente en la bahía de Molaix.

La película de petróleo extendida sobre las aguas representa una barrera para los microorganismos planctónicos, aparte del carácter tóxico de los hidrocarburos. Pero en las zonas mejilloneras se restablece, en un tiempo relativamente corto, el equilibrio roto por la presencia de partículas en suspensión, dado que, tras la filtración de las mismas, se eliminan pseudoheces por parte de los mejillones y otros bivalvos, sediméntandose en los fondos marinos, trasladándose el problema del plancton al bentos. Cabe señalar que tan nefastas son las alteraciones provocadas por los hidrocarburos del petróleo como las provocadas por los detergentes utilizados para su eliminación.

Bajo un aspecto no morfológico la incorporación de metales por parte de los mejillones ha preocupado a diversos autores. Así en 1972 Establier, del Laboratorio de Investigaciones Pesqueras de Cádiz, publicó en *Informes Técnicos* de dicho centro los resultados hallados en mejillones del noroeste de España del nivel de mercurio acumulado en estos lamelibranquios. Tras una homogeinización de los ejemplares, se someten a un espectrofotómetro de absorción atómica, encontrando concentraciones comprendidas entre las 0,06 a las 0,14 partes por millón, habiendo algunos lotes en que la cantidad de mercurio asciende de 0,24 a las 0,28 ppm, concentraciones por debajo de los límites tolerados en otros países.

Recientemente, la revista *Environmental Conservation* (vol. 5., páginas 1-25, 1978) ha publicado los resultados de una interesante labor llevada a cabo por Goldberg y sus colaboradores, quienes en 1975 iniciaron el estudio de lotes de mejillones procedentes de 107 estaciones costeras del este y oeste de Estados Unidos. Habiendo localizado no sólo metales pesados (hierro, mercurio, zinc, etcétera), sino también hidrocarburos del petróleo, hidrocarburos halogenados, núcleos transuránicos, así como productos resultantes de la degradación de insecticidas y herbicidas, habiéndose detectado la presencia de los dos principales productos resultantes de la degradación del DDT: p.p'DDE y p.p'DDD.

Una breve reseña de este estudio sobre el "mejillón-contador" ha aparecido en una nota de Bayne, profesor norteamericano interesado desde hace años en el ciclo biológico y en la productividad de los mejillones, publicada en la revista *Nature* (vol. 275, del 14 de septiembre de 1978).

Por todo ello podemos considerar que el mejillón, además de su interés como alimento de gran consumo en fresco y en conserva, por sus propiedades organolépticas y nutritivas (recuérdese la gran cantidad de glucógeno que tiene), presenta a la vez un gran interés bioecológico como lo demuestra la atención prestada en los últimos años por los científicos norteamericanos de la Scripps Institution of Oceanography, encabezados por Goldberg.

En España, desde principios de siglo, el mejillón ha merecido la atención como animal de consumo y su cultivo consiguiente, considerado por Ramón Margalef como una explotación modélica desde el punto de vista ecológico. El cultivo se inició en las costas de Tarragona en 1901, sobre cuerdas colgantes, pasando en 1909 a cultivarse en Barcelona. Esta modalidad de cultivo sobre bateas de cuerdas flotantes no llegó a Galicia hasta 1946, según datos publicados por Andréu, actual director del Instituto de Investigaciones Pesqueras de Barcelona.

De un trabajo publicado por los doctores López Benito, Gallardo y Gil, del Laboratorio de Investigaciones Pesqueras de Cádiz, sacamos los datos siguientes: en 1962 existían en Galicia 1327 mejilloneras, pasando a 2000 en 1963. En 1973, el número de bateas flotantes ascendía a 3300 con un total de dos millones trescientas mil cuerdas y una producción del orden de las 138.600 toneladas métricas. [La mayor parte de los parques flotantes de mejillones están en las rías bajas, siendo Arosa la que cuenta con mayor número, 2056 bateas frente a 507 instaladas en Vigo (datos de agosto de 1975)]. Lo cual indica el gran rendimiento económico que representa el cultivo del mejillón. De esta producción total de mejillones, el 50 por ciento se consume en fresco y el 50 por ciento restante en conservas de aceite y de escabeche.

Parece ser que el número actual se ha estabilizado, dadas las grandes dificultades que en los últimos años han tenido lugar en las costas gallegas: vertido de detergentes y contaminantes, presencia de petróleo

en las aguas, aparición de la marea roja, etcétera; tras una rotura del equilibrio por un proceso de este tipo las mejilloneras tardan unos siete años en volverse a estabilizar.

En España continúa siendo el mejillón un animal interesante por sus dos facetas: elemento de consumo y sensible indicador de contaminantes. (M. D.)

Nobel de medicina 1978

En el tiempo que el hombre ha practicado la agricultura y la ganadería, ha ido logrando especies más beneficiosas mediante técnicas ingenuas de selección con base en la recombinación de los genes o determinantes de caracteres. Ahora, 25 años después de que se estableciera la estructura en doble hélice del ADN, se dispone de una tecnología biológica que bien pudiera abrir una nueva era de domesticación de organismos. Tres adelantados en el alumbramiento de esta nueva era: los doctores Werner Arber, de la Universidad de Basilea, y Daniel Nathans y Hamilton O. Smith, de la Universidad John Hopkins de Baltimore, acaban de ser galardonados con el premio Nobel de medicina.

Cuando coinciden dos móleculas diferentes de ADN en una misma célula, puede ocurrir el fenómeno de recombinación. Las dos moléculas lineales, por acción de los enzimas de la célula, se pueden cortar y recomponer pero de modo que el tramo segundo de cada una pase a estar unido al tramo primero de la otra. Cuando esto ocurre, las dos nuevas moléculas son recombinantes de las originales: la descendencia las recibirá como legado. Sólo serán viables los descendientes cuando en la recombinación hayan intervenido dos moléculas de ADN muy semejantes entre sí y los cortes que hacen posible el intercambio de las porciones se hayan producido a la misma altura en ambas moléculas. De otra forma no conservarían una colección de genes suficiente para garantizar su supervivencia. De ahí que la recombinación *in vivo* sólo se observe entre mutantes de una misma especie o entre especies muy afines. Con la nueva tecnología bioquímica, la recombinación se realiza en el tubo de ensayo, entre moléculas que no tienen por qué parecerse. Se pueden obtener recombinantes que incluso procedan de una ADN eucariótico y un ADN procariótico. Se pueden crear quimeras de animal y bacteria.

Con las nuevas técnicas se puede hacer ingeniería genética basada en un conocimiento detallado con posibilidad de insertar unos pocos genes dentro de una molécula de ADN, como se puede intercalar una secuencia en un film, cortando la película y empalmando en el lugar preciso. Luego, la molécula recombinante artificial que contiene una porción que actúa de vehículo por transducción acaba por incorporar nuevos genes mediante un proceso natural en la dotación genética de la célula escogida. La molécula de ADN que actúa de vehículo contiene, además, un marcador: un gen que permita seleccionar, de entre las células, aquellas que han recibido e incorporado los nuevos genes. Estas células seleccionadas se pueden cultivar, a su vez, de forma que cada una genere una colonia y los genes introducidos se multipliquen.

Los supuestos riesgos de estas técnicas, que pudieran conducir a situaciones límite o crisis biológicas, han provocado muchos debates y han determinado la adopción de medidas racionales de precaución. Pero no cabe duda que estas mismas técnicas pueden conducir a desarrollos, sobre todo en agricultura (nuevas combinaciones de organismos capaces de fijar nitrógeno y de llevar a cabo fotosíntesis de alto rendimiento) y en medicina (producción económica de hormonas polipeptídicas y sustancias con actividad biológica) de los que bien pudiera depender, llegado el caso, la propia supervivencia de la humanidad civilizada.

A comienzos de la década de los años 50, Zinder y Lederberg descubren el fenómeno de transducción. Morese estudia por vez primera el fago λ en el laboratorio de los Lederberg en 1954. Juntos observan que las bacterias transformadas por transducción con fago λ casi siempre son inmunes al fago y son portadoras de profago defectivo. También trabajan sobre transducción por fago λ Kellenberger, Weigle y Arber. Por sugerencia de Weigle, Werner Arber comienza a trabajar sobre profagos defectivos como tema de su tesis doctoral que presenta en la Universida de Ginebra. A través de estos estudios, Arber logró aclarar el fenómeno que Luria había detectado a comienzos de la década y había descrito como variación inducida por el hospedador.

El fenómeno observado por Luria tiene relación con el cambio de especificidad en la relación entre fago y hospedador. Un fago que crece bien sobre una cepa bacteriana, crece mal sobre otra. Sin embargo, los pocos fagos que, a pesar de todo, logran crecer en la segunda presentan un curioso comportamiento: se desarrollan bien en la segunda cepa y mal en la primera. Arber, junto con Dussoix, explica este fenómeno de variación a través de una serie de experimentos netamente genéticos. En 1926, ambos proponen la existencia de un mecanismo de restricción-modificación controlado por el hospedador que, con el tiempo, se demostraría que era correcto.

La bacteria posee un juego de dos actividades enzimáticas: sistemas *R-M*, que reconocen una misma secuencia en un tramo corto de ADN. El enzima de restricción escinde la molécula de ADN en un punto del tramo si está sin modificar. El enzima de modificación altera, generalmente por metilación, una base de uno de los nucleótidos comprendidos en el tramo. Como se ve, la acción de ambos enzimas se excluye mutuamente. El ADN de la bacteria no se escinde porque está metilado en el lugar apropiado. El ADN de una fago que tenga acceso a la célula, pero que se reconoce como foráneo y se rechaza, mantiene la secuencia del tramo sin modificar y se degrada.

Después de proponer el mecanismo, Arber trabajó para establecer los detalles bioquímicos. En 1968 presentaba sus resultados sobre el enzima de restricción, que luego se denominaría *Eco B*. Aunque el enzima estudiado pertenece a la Clase I de enzimas de restricción, que tiene poco valor en genética bioquímica, sus trabajos hicieron posible la investigación de los enzimas realmente trascendentales de la Clase II. Hamilton O. Smith, junto con R. W. Wilcox, encontraron, en 1970, el enzima denominado Hind II en *Haemophilus influenzae*, primero de este tipo revolucionario en lo referente al capítulo de sus aplicaciones.

Los enzimas de restricción de la Clase I y de la Clase II reconocen un tramo con una secuencia característica de ADN. Los enzimas de la Clase I cortan ambas hebras del dúplex de ADN en varios puntos al azar y a menudo alejados de la secuencia reconocida: los productos de su acción sobre una molécula de ADN dada son variados. Los enzimas de la Clase II, en cambio, reconocen un tramo característico que, para todos los enzimas de esta clase, tiene la propiedad de que se lee igual en una hebra que en la otra, y cortan una y otra hebra en el mismo punto de las dos secuencias idénticas.

La determinación del orden en que se hallan los aminoácidos de una proteína dependió en gran medida de la existencia de enzimas digestivos, que son capaces de cortar la cadena polipeptídica junto a determinado tipo de residuo. La análoga determinación del orden en que estan dispuestos los nucleótidos en el ADN resultó muy laboriosa mientras sólo se dispuso de enzimas capaces de cortar atendiendo a la naturaleza de los nucleótidos, ya que estos son de naturaleza muy poco variada salvo en un tipo de ARN, el

de transferencia. La situación cambió por completo al aparecer los enzimas de restricción de la Clase II, como advirtió Daniel Nathan. Pero estos enzimas no cortan nucleótidos sino pequeñas secuencias de nucleótidos, con lo que su acción resulta muy específica y el número de cortes que producen en una molécula de ADN es muy pequeño. La aplicación de los enzimas de restricción de la Clase II les ha conferido, en la determinación de secuencias de ADN, el mismo papel que han desempeñado la tripsina y la quimotripsina en la determinación de secuencias de proteínas.

Cuando un enzima de restricción de la Clase II actúa sobre dos moléculas diferentes de ADN que poseen un mismo tramo que el enzima reconoce, generan nuevos extremos (en cada corte) que poseen secuencias limitadas idénticas en los productos que proceden de una y otra molécula. Además, muchos enzimas de esta Clase cortan las dos hebras de ADN en puntos desfasados, aunque simétricos, de la secuencia del tramo reconocido. En este caso, en todos los extremos creados por la acción enzimática, una de las hebras es más larga que la otra, es decir, en cada extremo existe una porción de la molécula de cadena sencilla, no doble. Esta porción tiene la misma secuencia, y esta secuencia es auto-complementaria. Ello hace que si el enzima actúa sobre dos moléculas de ADN diferentes genere en ambas extremos que se adhieren, facilitando asi la creación, a partir de ellas, de nuevas moléculas recombinantes.

La utilización de enzimas de restricción ha permitido ya logros tan espectaculares como el consistente en introducir el gen para la síntesis de la hormona somatostatina en *E. coli*, que se puede considerar como ejemplo de introducción de un gen eucariótico en un organismo procariótico, en el que además se expresa. Las experiencias realizadas con bacterias como células receptoras también deben ser factibles con células eucarióticas. Estas experiencias requieren vehículos apropiados que permitan implantar los genes en las células, además de un conocimiento claro de los complejos sistemas de regulación presentes en los organismos superiores. En el mismo Departamento de Smith, de la Universidad John Hopkins (Estados Unidos), comenzó a trabajar Daniel Nathans con SV40, que es un virus que ataca los simios produciendo tumores. Sus propias investigaciones acometidas durante los últimos años sobre la aplicación de enzimas de restricción a estos sistemas de tanta repercusión médica son los que le avalaron para compartir del premio con Arber y Smith. (E. C.).

El último teorema de Fermat

Durante 300 años los matemáticos se han esforzado en vano por demostrar un teorema que Fermat afirmaba poder probar: Ninguna potencia n-ésima puede resultar ser suma de otras dos potencias n-ésimas cuando n es mayor que 2

Harold M. Edwards

Pierre de Fermat, matemático francés del siglo XVII, fue fundador de la moderna teoría de números, rama de las matemáticas que estudia las propiedades de los números enteros. Lo mismo que tantos otros eruditos de su tiempo, estudió profundamente las obras clásicas de la antigüedad. En teoría de números, su fuente de inspiración fue Diofanto, matemático griego cuya *Aritmética* fue descubierta por los europeos a mediados del siglo XVI. Fermat comentó su ejemplar de esta obra con numerosas notas al margen, y tras su muerte, acaecida en 1665, su hijo publicó una nueva edición de la *Aritmética*, esta vez, con las anotaciones de su padre. Una de estas notas ha llegado a ser uno de los más famosos enunciados de la historia de las matemáticas.

Al lado de un problema relativo a determinar cuadrados expresables como suma de otros dos cuadrados (por ejemplo, 25 igual a 9 más 16), Fermat escribió (traducimos del latín): "Por otra parte, es imposible que un cubo sea suma de otros dos cubos, una cuarta potencia, suma de dos cuartas potencias, o en general, que ningún número que sea potencia mayor que la segunda pueda ser suma de dos potencias semejantes. He descubierto una demostración verdaderamente maravillosa de esta proposición que este margen es demasiado estrecho para contener". Tal proposición ha llegado a llamarse "último teorema" o "teorema magno" de Fermat. A pesar de que los más sagaces matemáticos lo han intentado en vano durante los tres siglos transcurridos, no se ha conseguido ninguna demostración válida. El último teorema de Fermat sigue siendo uno de los grandes problemas no resueltos de la matemática de nuestros días.

¿Tenía verdaderamente Fermat la "maravillosa demostración" que afirmaba? Era, sin duda, prodigioso matemático, que contribuyó a establecer las disciplinas hoy llamadas geometría analítica (con Descartes), cálculo diferencial (con Leibniz y Newton) y teoría de probabilidad (con Pascal). Su profesión no era la de matemático, sino jurista, y vivía en Toulouse, en la Francia meridional. Su amplia participación en la vida intelectual de su época se produjo enteramente mediante correspondencia privada con otros sabios de los principales centros europeos del saber.

Imaginar que un jurista provinciano del siglo XVII haya podido burlar con su teorema a los más capaces matemáticos de tres siglos resulta ciertamente encantador; pero los hechos parecen indicar que Fermat no disponía de verdadera demostración. A excepción de la nota escrita al margen del libro de Diofanto, ninguno de los escritos de Fermat que han sobrevivido hace mención alguna sobre la demostración del teorema. Sí se menciona en algún otro lugar que sabía cómo probar la imposibilidad de soluciones para las ecuaciones $x^3 + y^3 = z^3$ y $x^4 + y^4 = z^4$. Si Fermat hubiera descubierto una demostración válida del teorema general (a saber, que es imposible que $x^n + y^n = z^n$, siendo x, y, z y n enteros positivos, y n mayor que 2) sería sorprendente que no la hubiera mencionado también. Lo más verosímil parece ser que, en el momento de escribir la nota, Fermat tenía una idea de cómo demostrarla, que más tarde resultó insuficiente. Puede decirse con casi absoluta certeza que al escribir sus notas Fermat no tenía el propósito de publicarlas, y pudiera ser que no tuviera ocasión de volver atrás y borrar o enmendar la citada.

Desde luego, resulta mucho más emocionante pensar que Fermat sí disponía de una demostración rigurosa del teorema, y hasta es posible que así sucediera. En cualquier caso, el teorema ha tenido gran influencia en el desarrollo de la teoría de números. Algunas de las máximas creaciones del pensamiento matemático han sido sugeridas por su estudio, y las técnicas desarrolladas en el esfuerzo de demostrarlo han contribuido a la solución de otros muchos problemas.

Finalmente, la historia del último teorema de Fermat proporciona excelente ilustración de la verdadera naturaleza de la indagación matemática. Suele preguntarse a los matemáticos: "Pero, ¿cómo es posible investigar en matemáticas?" Dudo mucho que llegue a planteársele a un físico o biólogo semejante pregunta. Mucha gente está convencida de que las matemáticas son tema tan rutinario que en ellas el trabajo apenas puede ir más allá de la ordenada consignación de los datos. Evidentemente, nada más lejos de la realidad. En matemáticas, como en cualquier otro terreno, hay cuestiones no resueltas por doquier; para los matemáticos, la dificultad reside en dar con preguntas que puedan contestar, y no al contrario. No es fácil, sin embargo, darle al lego ejemplos claros que ilustren este punto, porque el enunciado de las cuestiones de interés matemático suele requerir terminología especializada y formación suficiente. El teorema magno de Fermat es rara excepción a esta regla.

Demostración por descenso infinito

Parte de la fascinación del último teorema de Fermat procede del hecho de ser tan sencillo de enunciar y comprender: Es imposible hallar números enteros positivos x, y, z y n, donde n sea mayor que 2 y se verifique la igualdad $x^n + y^n = z^n$. Los

PIERRE DE FERMAT, llamado "padre de la teoría de números", rama de las matemáticas que estudia las propiedades de los números enteros. Nacido en 1601, cerca de Toulouse, Fermat pasó toda su vida en el sur de Francia, lejos de los grandes centros europeos del saber. No era matemático profesional, sino jurista, y ninguno de sus trabajos de matemáticas vio la luz pública hasta después de su muerte. Su amplia participación en las matemáticas de su tiempo se realizó por completo a través de correspondencia particular con otros estudiosos. Fermat enunció muchos estimulantes, profundos y difíciles teoremas, que no fueron demostrados hasta mucho después de su muerte. Hacia 1840, solamente faltaba demostrar uno de ellos, que ha llegado a denominarse "último teorema de Fermat": Cuando n es un entero mayor que 2, no existe ninguna solución de la ecuación $x^n + y^n = z^n$ formada exclusivamente por números enteros. El último teorema de Fermat es uno de los más famosos problemas pendientes de solución de la matemática. El retrato se halla en la Académie des Sciences, Inscriptions et Belles Lettres de Toulouse; se reproduce con permiso de R. Gillis.





UNA TABLETA BABILONICA de arcilla, grabada con caracteres cuneiformes, que data aproximadamente de 1500 a.C., y es uno de los más antiguos documentos sobre teoría de números. La tableta recopila (aunque de forma ligeramente encubierta) varios conjuntos de ternas pitagóricas, esto es, ternas de números enteros positivos, x, y, z tales que $x^2 + y^2 = z^2$, como, por ejemplo, 4961, 6480 y 8161. Fermat enunció su último teorema mientras analizaba un problema referente a ternas pitagóricas *(véase la ilustración de la página siguiente).* Cualquiera de estas ternas prueba que su teorema es falso cuando el exponente n sea igual a 2. Las ternas pitagóricas reciben dicho nombre de su relación con el teorema de Pitágoras, que enuncia que en todo triángulo rectángulo el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los otros dos lados. No es fácil encontrar ternas por simple tanteo, sobre todo cuando los números son grandes, por lo que seguramente los babilonios tenían algún método para hallarlas. Lo más probable parece ser que su interés por ellas proviniera de las aplicaciones geométricas de estos conjuntos de números, y que, por tanto, conocieran ya el contenido del teorema de Pitágoras 1000 años antes que los griegos. La tableta forma parte de la colección Plimpton de la Biblioteca Butler, de la Universidad de Columbia, por cuya gentileza la reproducimos.

no iniciados en matemáticas suelen atacar el teorema por el método en apariencia más razonable: ensayando valores. Tomemos el caso de que n sea igual a 3, a saber, el caso consistente en demostrar que la ecuación $x^3 + y^3 = z^3$ no tiene ninguna solución. Los cubos de los 10 primeros números enteros positivos son 1, 8, 27, 64, 125, 216, 343, 512, 729 y 1000. No es difícil ver que ninguno de estos números puede descomponerse en suma de otros dos cubos. Por ejemplo, si 512 fuese suma de otros dos cubos, éstos serían menores que 512, y figurarían antes que él en la lista. Los cubos 216 y 343 son buenos candidatos, pero su suma es 559, demasiado grande. La suma inmediatamente menor, 216 más 216, es igual a 432, y no 512. Así pues, 512 no es suma de ningún par de cubos.

Es sencillo comprobar por este procedimiento que un cubo concreto dado no es suma de ningún par de cubos. Los cálculos requeridos podrían efectuarse a gran velocidad por un ordenador, siendo de esta forma fácil demostrar que ningún cubo de, pongamos por caso, menos de diez dígitos es suma de otros dos cubos. Sin embargo, hay una infinidad de cubos a comprobar, y así pues, por este procedimiento ni aún con auxilio del más potente y rápido ordenador podrá resolverse la cuestión de saber si algún cubo podrá descomponerse en suma de otros dos.

Una vez que el lego en matemáticas se ha convencido de que los ensayos no bastan para demostrar la imposibilidad de $x^3 + y^3 = z^3$ suelen irse al otro extremo, y poner en tela de juicio que nadie pueda probar un enunciado semejante. La respuesta es que podría probarse por reducción al absurdo: se supone inicialmente que la ecuación sí tiene solución, y de esta hipótesis es necesario deducir un enunciado cuya falsedad sea ya conocida. De llegarse a un enunciado falso —contradicción o absurdo— estaría demostrada la falsedad de la suposición inicial, lo que significaría que no puede existir ninguna solución.

Más concretamente, en el caso de enunciados referentes a enteros positivos, como es evidente que ocurre en el último teorema de Fermat, las demostraciones por reducción al absurdo suelen adoptar la forma de demostraciones por descenso infinito. Fermat mantuvo haber sido el creador del método, que decía era base de todas sus demostraciones en teoría de números. En una demostración por descenso infinito hay que probar que de existir una solución formada por enteros positivos para la ecuación que nos ocupa, a partir de ella puede construirse otra solución formada por enteros positivos más pequeños. El mismo razonamiento muestra entonces que la segunda solución permite construir una tercera todavía más pequeña; el proceso puede reiterarse indefinidamente. Es-

tando todas las soluciones formadas por enteros positivos, es evidentemente imposible hallar soluciones cada vez más pequeñas *ad infinitum*. Por consiguiente, no puede existir solución alguna.

Primeros descubrimientos

No hay en todos los escritos de Fermat, sobre la teoría de números, que hayan llegado hasta nuestros días más que una sola demostración, descubierta en otra nota marginal a la *Aritmética* de Diofanto. Se refiere a triángulos pitagóricos, es decir, triángulos rectángulos cuyos lados tienen longitudes enteras. (El nombre procede de su relación con el teorema de Pitágoras, que establece que los catetos x e y, y la hipotenusa z de un triángulo rectángulo están ligados por la relación $x^2 + y^2 = z^2$.) Fermat demostró que el área de un triángulo pitagórico no puede ser cuadrado perfecto de ningún número entero: Si x, y y z son enteros positivos tales que $x^2 + y^2 = z^2$, entonces $(1/2)xy$ no es cuadrado de ningún entero. No es difícil demostrar que alguno de los números x o y tiene que ser par, y que por consiguiente $(1/2)xy$ es entero.

Para la demostración, Fermat utilizó el método de descenso infinito. Concretamente, dio un procedimiento explícito mediante el cual, dados enteros x, y, z y u que cumplan las condiciones $x^2 + y^2 = z^2$ y $(1/2)xy = u^2$, es posible construir otro conjunto de números enteros X, Y, Z y U tales que $X^2 + Y^2 = Z^2$, que $(1/2)XY = U^2$, y que el triángulo de lados X, Y, Z sea más pequeño que el de lados x, y, z en el sentido de que la hipotenusa Z sea más pequeña que la hipotenusa z. El método utilizado por Fermat para construir X, Y, Z y U es francamente sutil, y requeriría prolija explicación. No obstante, la existencia de tal método muestra que es imposible hallar soluciones enteras para el sistema de ecuaciones $x^2 + y^2 = z^2$; $(1/2)xy = u^2$, pues una tal solución conllevaría otra solución más pequeña X, Y, Z y U, que a su vez daría una solución X', Y', Z' y U' todavía menor, y así sucesivamente. Se deduce entonces que existiría una sucesión infinitamente decreciente de números enteros positivos $z > Z > Z' >$, lo cual es imposible. (El signo "$>$" significa "mayor que".)

Hecho notable, y probablemente nada casual, es que la imposibilidad de ser cuadrado perfecto el área de un triángulo pitagórico implique inmediatamente que la ecuación $x^4 + y^4 = z^4$ carezca de solución, es decir, que el teorema de Fermat sea verdadero cuando n es igual a 4. Un sencillo e ingenioso artificio permite conectar ambas proposiciones. Supongamos que $x^4 + y^4 = z^4$ para ciertos enteros positivos x, y, z y hagamos y^4 igual a a, $2x^2z^2$ igual a b, $z^4 + x^4$ igual a c y, finalmente, y^2xz igual a d. Entonces, la sencilla identidad algebraica $(r + s)^2 = r^2 + 2rs + s^2$ implica que $a^2 + b^2$ sea igual a $(z^4 - x^4)^2 + 4x^4z^4$, o lo que es igual, a $z^8 - 2x^4z^4 + x^8 + 4x^4z^4$, o sea, a $(z^4 + x^4)^2$, que es igual a c^2. Además, $(1/2)ab$ es igual a $(1/2)y^4 2x^2z^2$, o sea, igual a $(y^2xz)^2$, que es igual a d^2. Por consiguiente, $a^2 + b^2 = c^2$ y $(1/2)ab = d^2$, que Fermat demostró ser imposible. Por consiguiente, la hipótesis de partida, que $x^4 + y^4 = z^4$ tenía al menos una solución, ha de ser incorrecta, con lo que el teorema de Fermat queda demostrado en el caso de n igual a 4. Así pues, en esencia, el propio Fermat demostró su teorema para el caso de cuartas potencias.

DIOPHANTI
ALEXANDRINI
ARITHMETICORVM
LIBRI SEX,
ET DE NVMERIS MVLTANGVLIS
LIBER VNVS.
CVM COMMENTARIIS C. G. BACHETI V. C.
& obseruationibus D. P. de FERMAT *Senatoris Tolosani.*
Accessit Doctrinæ Analyticæ inuentum nouum, collectum
ex varijs eiusdem D. de FERMAT Epistolis.



TOLOSÆ,
Excudebat BERNARDVS BOSC, è Regione Collegij Societatis Iesu.
M. DC. LXX.

EL ULTIMO TEOREMA DE FERMAT fue consignado por primera vez en el margen de la *Aritmética*, obra dedicada a teoría de números cuyo autor fue Diofanto, matemático de la Grecia clásica. Fermat estudió cuidadosamente el libro, e hizo numerosas anotaciones marginales en su ejemplar de esta obra, una traducción latina debida a C. G. Bachet. Tras la muerte de Fermat, en 1665, su hijo publicó una segunda edición de la traducción de Bachet de *Aritmética* que incluía en un apéndice las notas marginales de Fermat. En esta ilustración se muestra la portada del libro. Traduciendo del latín, anuncia *Seis libros de aritmética y un libro sobre números poligonales, por Diofanto de Alejandría*, "con comentarios del distinguido caballero C. G. Bachet y observaciones del Señor P. de Fermat, Senador de Tolosa" y "un nuevo descubrimiento de Doctrina Analítica, recopilada de diversas cartas del mismo señor de Fermat". La histórica nota de Fermat se refería a un problema relativo a ternas pitagóricas. He aquí sus palabras: "Por otra parte, es imposible... que ningún número que sea potencia mayor que la segunda pueda ser suma de dos potencias semejantes. He descubierto una demostración verdaderamente maravillosa de esta proposición, que este margen es demasiado estrecho para contener". Se ignora si Fermat tenía una demostración rigurosa del teorema. En cualquier caso, ningún matemático ha podido ni demostrar el teorema ni hallar un contraejemplo que lo invalide.

La demostración anterior establece también que el teorema de Fermat es verdadero siempre que n sea múltiplo de 4, porque si n es igual a $4k$ para algún entero positivo k, entonces $x^n + y^n = z^n$ implica $(x^k)^4 + (y^k)^4 = (z^k)^4$, lo que es imposible, ya que ninguna cuarta potencia puede ser suma de otras dos cuartas potencias. Exactamente de la misma forma, si el teorema puede ser demostrado para un exponente determinado, m, entonces el teorema es verdadero para todos los múltiplos de m. Así pues, dado que todo entero n que sea mayor que 2 es divisible bien por 4, bien por un número primo impar (es decir, por un número primo distinto de 2), bastará demostrar el teorema de Fermat en todos los casos en que el exponente sea un número primo impar.

Fermat aseguró poder demostrar el teorema en el caso de ser n igual a 3; pero no se publicó ninguna demostración de la imposibilidad de que hayan soluciones de $x^3 + y^3 = z^3$ hasta cien años más tarde. Tal demostración, obra del matemático suizo Leonhard Euler, contenía, sin embargo, un grave fallo.

La demostración de Euler

Euler dio una demostración por descenso infinito, consistente en un método para deducir de cada solución x, y, z de la ecuación $x^3 + y^3 = z^3$ una nueva solución X, Y, Z donde Z sea menor que z. Su método es demasiado largo para poder exponerlo aquí con detalle, pero a grandes rasgos, efectuando cálculos relativos a las diversas características de x, y, z, Euler redujo el problema de deducir una solución más pequeña al de probar la siguiente proposición: Si p y q son números enteros primos entre sí (es decir, que no tienen más divisores comunes que la unidad) y si $p^2 + 3q^2$ es cubo perfecto, entonces han de existir enteros a y b tales que p sea igual a $a^3 - 9ab^2$ y que q sea igual a $3ab^2b - 3b^2$. Tal proposición es perfectamente correcta, y puede demostrarse aplicando variantes sencillas de métodos expuestos en diversos lugares de la obra publicada de Euler. En esta ocasión, sin embargo, Euler prefirió utilizar un nuevo tipo de razonamiento que requería introducir números de la forma $a + b\sqrt{-3}$, siendo a y b dos enteros.

Para comprender por qué Euler vio útiles los números de la forma $a + b\sqrt{-3}$, desarrollaremos la expresión $(a + b\sqrt{-3})^3$. Resulta ser $a^3 + 3a^2 b\sqrt{-3} - 9ab^2 - 3b^3\sqrt{-3}$, o sea $(a^3 - 9ab^2) + (3a^2b - 3b^3)\sqrt{-3}$, es decir, $p + q\sqrt{-3}$, donde p y q son números definidos como en la conclusión de la proposición. Dicho de otra forma, la tesis de la proposición afirma que $(a + b\sqrt{-3})^3 = p + q\sqrt{-3}$. Ahora bien, la proposición hace la hipótesis de que $p^2 + 3q^2$ es cubo perfecto. Volviendo a escribir $p^2 + 3q^2$ en la forma $(p + q\sqrt{-3})(p - q\sqrt{-3})$, la proposición puede ahora enunciarse así: Si $(p + q\sqrt{-3})(p - q\sqrt{-3})$ es cubo perfecto, entonces $p + q\sqrt{-3}$ ha de ser un cubo, es decir, ha de ser de la forma $(a + b\sqrt{-3})^3$, para ciertos enteros a y b. Por consiguiente, al introducir los números $a + b\sqrt{-3}$ la proposición puede enunciarse de forma mucho más sencilla y natural.

Los números de la forma $a + b\sqrt{-3}$ constituyen un sistema numérico que comparte muchas propiedades de los enteros. En ambos casos, al sumar, restar o multiplicar dos números cualesquiera del sistema resulta otro número perteneciente al sistema, aunque por lo general no ocurre así al dividir. Por ejemplo 5 no es divisible entre 4 en el sistema de los números enteros (es decir, no hay ningún entero que multiplicado por 4 dé 5), y dados dos números de la forma $a + b\sqrt{-3}$, por lo general tampoco es posible hallar un tercer número de esa misma forma que sea su cociente. Las semejanzas de ambos sistemas indujeron a Euler a dar un paso innovador —e incorrecto— en su demostración. Euler aplicó una propiedad válida sólo para enteros a los números $a + b\sqrt{-3}$.

La propiedad de los enteros que Euler precisaba para su demostración es consecuencia de la unicidad de la descomposición de cada entero en un producto de potencias de factores primos. Cada entero positivo solamente es expresable como producto de factores primos de una sola forma. Por ejemplo, 124 es igual a $2 \times 2 \times 31$; ningún número primo, salvo 2 y 31, puede dividir exactamente a 124. El teorema de factorización única implica en los enteros la siguiente propiedad: La única forma de que el producto de dos



LEONHARD EULER demostró el teorema de Fermat para el caso del primer exponente primo e impar, n igual a 3. Este retrato, realizado alrededor de 1745, muestra al matemático suizo durante su vida. Antes del trabajo de Euler el último teorema de Fermat había sido establecido (por aplicación directa del propio trabajo de Fermat) en tan sólo un caso: n igual a 4. Es fácil demostrar que si el teorema es cierto cuando n es igual a un cierto entero r entonces es verdadero cuando n sea cualquier múltiplo de r. Como todo entero es, o bien divisible por 4 o bien divisible entre un número primo impar, solamente faltaba demostrar el teorema en los casos donde n sea un número primo e impar. El primer intento dirigido a este fin que tuvo éxito fue el de Euler, pero aunque su demostración era ingeniosa tenía un grave fallo. El fallo fue fácil de corregir en el caso de n igual a 3, pero fue piedra angular de muchos esfuerzos posteriores por intentar probar el teorema.

enteros primos entre sí sea un cubo perfecto es que cada factor sea cubo perfecto. Por ejemplo, supongamos que c y d sean números primos entre sí, y que cd sea igual a 1000, es decir, a 10^3. Expresado como producto de factores primos, 1000 es igual a $2^3 \times 5^3$. Las distintas descomposiciones factoriales de 1000 se obtienen todas distribuyendo sus factores primos en dos subconjuntos; tal, por ejemplo, $(2 \times 2 \times 5)$ $(2 \times 5 \times 5)$, o sea, 20×50. Si el reparto debe efectuarse de manera que los dos conjuntos sean primos entre sí, todos los factores 2 deberán ir a uno de los dos subconjuntos, y todos los factores 5 deberán ir al otro. Por consiguiente, los únicos valores posibles de los factores c y d son $2^3 \times 5^3$, 2^3, 5^3 y 1, que son todos ellos cubos perfectos. Análogamente, el producto de dos enteros primos entre sí solamente puede ser una potencia n-ésima cuando cada uno de ellos sea por sí mismo potencia n-ésima.

Euler supuso que esta propiedad de los enteros era poseída también por los números $a + b\sqrt{-3}$, y demostró su proposición razonando de la siguiente forma: La proposición afirma que si p y q son primos entre sí y el producto $(p + q\sqrt{-3})$ $(p-q\sqrt{-3})$ es cubo perfecto, entonces $p + q\sqrt{-3}$ es cubo perfecto. Euler demostró en primer lugar que si p y q son primos entre sí, entonces $p + q\sqrt{-3}$ y $p-q\sqrt{-3}$ son también primos entre sí. Por extensión de la propiedad de los enteros, el producto de números primos entre sí de la forma $a + b\sqrt{-3}$ solamente puede ser cubo perfecto si lo son los propios factores. Por tanto, la hipótesis de ser $(p + q\sqrt{-3})$ $(p-q\sqrt{-3})$ cubo perfecto implica que $p + q\sqrt{-3}$ es cubo perfecto, lo que demuestra la proposición.

El fallo del trabajo de Euler

Este razonamiento ciertamente lleva a Euler hasta la conclusión requerida; el problema está en que no basta razonar por analogía con los enteros para obtener una demostración válida. Los razonamientos por analogía son extraordinariamente sugestivos, y como demuestra la historia de las matemáticas, pueden ser origen de fecundas ideas, pero por sí solos no demuestran nada. Sorprende especialmente que Euler no mostrara mayor circunspección al usar este tipo de analogías, pues aunque los números $a + b\sqrt{-3}$ comparten muchas propiedades con los enteros, hay también muchos puntos de divergencia entre ambos sistemas. Por ejemplo, los enteros tienen una ordenación natural,...-2, -1, 0, 1, 2,... pero los números $a + b\sqrt{-3}$ no pueden tenerla. Sólo hay una forma de estar seguros de que los números $a + b\sqrt{-3}$ poseen, lo mismo que los enteros, la propiedad de que el producto de números primos entre sí solamente es cubo perfecto cuando lo sea cada uno de los factores, y es demostrar que así sucede.

No obstante, incluso los mejores matemáticos sucumben de cuando en cuando a la tentación de hacer demostraciones por analogía, descuidando criticar suficientemente ciertos razonamientos porque saben de antemano que la conclusión a que conducen es correcta. Esta tentación alcanza especial intensidad cuando el razonamiento tiene la seductora simplicidad que posee el de Euler. Es muy posible que Euler no tuviera mayor preocupación por el rigor de su demostración porque sabía ya por consideraciones de otra especie que la conclusión final $p + q\sqrt{-3} = (a + b\sqrt{-3})^3$ era correcta. Muchos años antes de publicar su demostración del caso n igual a 3 del último teorema de Fermat, Euler había trabajado ya en otras proposiciones enunciadas, pero no demostradas, por Fermat referentes a la representación de números en la forma $x^2 + 3y^2$. En particular, demostró el aserto de Fermat de que todo número primo p que sea uno más que un múltiplo de tres (p igual a $3n + 1$) tiene una única representación en forma de cuadrado más triple de un cuadrado (p igual a $x^2 + 3y^2$) por ejemplo, 7 es igual a $2 \times 3 + 1$ y a $2^2 + 3 \times 1^2$. Las técnicas desarrolladas por Euler para demostrar este teorema se aplican sin dificultad para demostrar la proposición del último teorema de Fermat en el caso n igual a 3. Es posible que Euler se diera cuenta de que podía demostrar la proposición con técnicas debidamente establecidas ya, y, por consiguiente, no sometiera la nueva y un tanto insólita demostración a escrutinio suficientemente cuidadoso.

En la otra demostración, anterior, Euler fue extraordinariamente cauto a la hora de utilizar razonamientos cuestionables. Por ejemplo, uno de los pasos intermedios de tal demostración requiere comprobar que si a y b son primos entre sí, entonces todo divisor primo e impar de $a^2 + 3b^2$ puede ser descompuesto en la forma $c^2 + 3d^2$. En esta ocasión, Euler no tenía la peligrosa certidumbre de que el enunciado a demostrar fuera correcto, y su demostración es modelo de claridad y rigor. A pesar de todo, no deja de llamar la atención que la experiencia adquirida en esta demostración no le previniera acerca de la escasa solidez de su segundo razonamiento. El razonamiento por analogía puede utilizarse para demostrar que todo divisor primo e impar de $a^2 + 3b^2$ es de forma $c^2 + 3d^2$, pero también podría usarse para probar que todo divisor primo e impar de $a^2 + 5b^2$ es de la forma $c^2 + 5d^2$, lo que es falso. (Observemos que $4^2 + 5 \times 1^2$ es igual a 21, y que ninguno de los divisores de 21, a saber, ni 3 ni 7, pueden expresarse como suma de un cuadrado más cinco veces un cuadrado.) Estas consideraciones han debido inspirar el característico rigor de la primera demostración de Euler. Sin duda, las había olvidado cuando volvió sobre el tema muchos años más tarde.

El lapsus de rigor de Euler está en estrecha relación con sus extraordinarias dotes, su inventiva e imaginación. Su capacidad para percibir nuevas conexiones entre diversos problemas, y para formularlos con originalidad y perspicacia hicieron de su obra fuente de inspiración de varias generaciones de matemáticos. La descomposición de los $a + b\sqrt{-3}$ en producto de factores primos era muestra clara de su ingenio extraordinario. De hecho, aunque la aplicación de esta idea al caso n igual a 3 del último teorema de Fermat fuese prematura, hechos posteriores mostraron que se trata de una idea llena de inspiración. Efectivamente, el fallo del razonamiento de Euler –a saber, que una propiedad deducida del teorema de factorización única de los enteros no tiene por qué ser válida forzosamente en otros sistemas de números semejantes a ellos– resultaría ser tema central de otras investigaciones de más altos vuelos sobre el teorema.

Dos casos más

Aunque la demostración de Euler precisaba de algunos remiendos, en esencia sí establecía la validez del último teorema de Fermat cuando n es igual al primer número impar, o sea, 3. En la década de 1820, Gustav Lejeune Dirichlet y Adrien-Marie Legendre demostraron el teorema para el caso de ser n igual a 5, que es el número primo sucesivo. Su método de demostración es, básicamente, generalización del utilizado por Euler para el caso n igual a 3; siendo en esta nueva situación numérica la ecuación $p + q\sqrt{5} = (a + b\sqrt{5})^5$ la análoga de la ecuación crucial $p + q\sqrt{-3} = (a + b\sqrt{-3})^3$. (Cuando n crece las ecuaciones que aparecen al intentar demostraciones de este estilo se complican demasiado, y el método ya no funciona.) En el caso n igual a 5, para demostrar que $p + q\sqrt{5}$ es una quinta potencia no solamente hay que suponer que p y q sean primos entre sí, como en el caso de ser n igual a 3, sino que además han de tener paridad contraria (es decir, ser uno par y el otro impar) y además, q tiene que ser divisible por 5. Para demostrar este hecho, Dirichlet hizo un estudio sumamente cuidadoso de los números de la forma x^2-5y^2. Su demostración, que tomaba por modelo otros trabajos de Eu-

ler, incluido su riguroso análisis de los números de la forma $x^2 + 3y^2$ así como otros trabajos de Joseph-Louis Lagrange y Carl Friedrich Gauss, no dependía de ninguna forma de analogías con las factorizaciones de los enteros.

Hubieron de transcurrir unos 15 años antes de que Gabriel Lamé demostrase el último teorema de Fermat para el caso del número primo sucesivo: 7. Su demostración fue un gran logro, pero no buen presagio, porque era larga, difícil y estaba fuertemente ligada al número 7. Había pocas esperanzas de poder aplicar tal demostración al caso siguiente, n igual a 11, o a cualquier otro caso del teorema. Se tenía entonces la impresión de que no se conseguirían progresos substanciales en tanto no se descubriera un nuevo enfoque del problema.



ERNST EDUARD KUMMER realizó los máximos progresos que se hayan conseguido nunca respecto al teorema de Fermat. En 1847, este especialista alemán de teoría de números, trabajando sobre un sistema de números llamados enteros ciclotómicos, descubrió una condición suficiente para que un entero primo p sea exponente de validez del teorema de Fermat, es decir, que si el número primo p verifica la condición de Kummer, entonces es imposible que $x^p + y^p = z^p$. Los números primos que satisfacen esta condición se llaman actualmente primos regulares. Cuando se descubrieron los primos regulares, el teorema de Fermat solamente estaba demostrado para los casos de n igual a 3, 4, 5 y 7; Kummer pudo demostrarlo para todos los números primos menores que 100, excepto tres (37, 59 y 67). Más recientemente se han descubierto otras condiciones suficientes mucho más generales que la de Kummer. Todos los números primos comprendidos en el margen de trabajo de las actuales técnicas de cálculo, incluso sobre los más grandes computadores, han resultado ser exponentes de validez del último teorema de Fermat. No obstante, el teorema no ha sido demostrado. En realidad, ni siquiera ha podido establecerse para un conjunto infinito de exponentes primos.

El propio Lamé propuso un tal método en 1847, al intentar demostrar el teorema general introduciendo las raíces n-ésimas de la unidad: una raíz n-ésima de la unidad es un número complejo α tal que α^n es igual a 1, y que α^k no es igual a 1 para ningún entero positivo k menor que n. Tal idea no era nueva. Ya en el siglo anterior Lagrange había señalado que introduciendo α en el estudio del último teorema de Fermat es posible descomponer $x^n + y^n = z^n$ en factores, cada uno de los cuales contiene x e y elevados a la primera potencia. (Por lo general, las potencias son tanto más sencillas de manejar cuanto más bajas sean.)

Para obtener tal descomposición basta notar que $1, a, a^2 \ldots a^{n-1}$ son las raíces, o soluciones de la ecuación $X^n - 1$; así en virtud del teorema fundamental del álgebra, $X^n - 1 = (X-1)(X-a)(X-a^2)\ldots(X-a^{n-1})$. Pongamos ahora X igual a $-x/y$ y multipliquemos ambos miembros de la ecuación por y^n. Como sólo se toman en cuenta los casos en que n sea impar, la ecuación resultante será $x^n + y^n = (x + y)(x + \alpha y)\ldots(x + \alpha^{n-1}y)$.

Los enteros ciclotómicos

Cada uno de los factores de $x^n + y^n$ es un número de la forma $a_0 + a_1\alpha + a_2\alpha^2 + \ldots + a_{n-1}\alpha^{n-1}$ donde $a_0, a_1, \ldots a_{n-1}$ son números enteros. En la actualidad los números de este tipo –números pertenecientes al sistema formado por enteros y potencias de α– se les llama enteros ciclotómicos, porque las raíces n-ésimas de la unidad están estrechamente emparentadas con el problema de división de una circunferencia en n partes iguales. (El número complejo α puede interpretarse como un punto situado sobre una circunferencia de radio unidad y cuyo centro sea el origen del plano complejo; el arco de la circunferencia comprendido entre 1 y α es $1/n$-ésimo de la circunferencia completa.) Lo mismo que los números $a + b\sqrt{-3}$, los enteros ciclotómicos constituyen un sistema semejante a los enteros, pues al sumar, restar o multiplicar enteros ciclotómicos resulta otro entero ciclotómico, aunque al dividirlos, por lo general, no es así.

El tratamiento dado por Lamé a los enteros ciclotómicos recuerda al de Euler para los números $a + b\sqrt{-3}$, aunque pudiera haber sido invención independiente. Dada la factorización de $x^n + y^n$ en enteros ciclotómicos, Lamé propuso que se aplicara el "hecho" de que el producto de números primos entre sí (aquí, al decir números se refería a los números enteros

ciclotómicos) solamente puede ser una n-ésima potencia si cada uno de dichos números es una n-ésima potencia. Llegado a este punto, Lamé vio abrirse ante sí un camino directo para terminar de probar que la igualdad $x^n + y^n = z^n$ es imposible. El problema es que una propiedad deducida de la unicidad de los enteros ordinarios no es más aplicable a los enteros ciclotómicos que a los números $a + b\sqrt{-3}$. Aunque en realidad esta propiedad sí es válida para los números de la forma $a + b\sqrt{-3}$ (si bien la razón no es evidente), no es en general verdadera para enteros ciclotómicos; con mayor precisión, sólo es verdadera para unos cuantos valores de n, donde n es un número primo impar y α es una raíz n-ésima de la unidad.

El desafortunado Lamé se vio tan arrastrado por su propio optimismo que anunció en una sesión de la Academia Francesa de Ciencias haber demostrado por su método el último teorema de Fermat. Sin embargo, apenas hubo presentado un esquema de su demostración, Joseph Liouville se levantó para objetar la forma en que Lamé aplicaba propiedades de los enteros ordinarios a la aritmética de los enteros ciclotómicos. No se sabe si Liouville tenía noticia del parecido error cometido por Euler. En cualquier caso, es notable que supiera detectar instantáneamente el punto más débil del razonamiento de Lamé. Había, además, otros puntos vulnerables. Tan exuberante era el entusiasmo de Lamé que éste pasó por alto otras varias dificultades serias. De hecho, su método resultó impracticable, ni siquiera para los valores excepcionales de n donde se verifica su hipótesis principal, a saber, n igual a 3, 5, 7, 11, 13, 17 y 19.

Como es natural, Lamé quedo avergonzado por sus simplezas, más aún por ser publicadas en las Actas de la Academia Francesa con lo que sus errores quedaban a la vista de todo el mundo matemático. "Si hubieras estado en París, o yo en Berlín", le escribió Lamé a su amigo Dirichlet, en Berlín, "nada de esto hubiera sucedido." En realidad, habría bastado que Lamé hubiese leído las Actas de la Academia de Ciencias de Berlín, donde pocos meses antes había aparecido el anuncio de una nueva e importante teoría sobre la aritmética de los enteros ciclotómicos.

La contribución de Kummer

El creador de la nueva teoría era Ernst Eduard Kummer. Hacía varios años que Kummer se había dado cuenta de que en problemas de teoría de números, como el último teorema de Fermat, la propiedad fundamental de los enteros ordinarios es la unicidad de la descomposición en producto de factores primos, y, por tanto, intentó demostrar esta propiedad para los enteros ciclotómicos. Lo que demostró, en cambio, es que en general tal propiedad no es válida para dichos números. (Había comunicado este descubrimiento en 1844, pero apareció en una publicación casi desconocida.) Conforme Kummer continuaba trabajando sobre los enteros ciclotómicos, fue viendo cada vez con mayor claridad que el teorema de factorización única, no válido en general para los enteros ciclotómicos, tampoco se necesitaba en toda su fuerza. Su teoría de 1847 ponía de manifiesto que el concepto de factorización única podía modificarse de manera que en su nueva versión pudiera aplicarse a la demostración de ciertas propiedades sutiles y útiles de los enteros ciclotómicos.

La base de la teoría de Kummer consistía en introducir en la aritmética de enteros ciclotómicos lo que llamó factores primeros ideales, de forma en algunos aspectos análoga a la introducción del número i, o sea, $\sqrt{-1}$, en la aritmética de los números enteros ordinarios. No examinaré el carácter ni las propiedades de los números ideales de Kummer, salvo para decir que restituyeron a los enteros ciclotómicos y a otros sistemas de números, como los de la forma $a + b\sqrt{-3}$, algunas de las más importantes consecuencias del teorema de factorización única, necesarias para probar el último teorema de Fermat. La teoría de la factorización por ideales fue sin duda uno de los más altos logros de la matemática decimonónica. La terminología ha evolucionado de forma extravagante y caprichosa, y los números ideales de Kummer, así como ciertos conjuntos de números asociados a ellos, han terminado por llamarse "ideales". En la actualidad, la teoría de ideales es por sí sola una rama de las matemáticas, nuevo testimonio de la extraordinaria importancia de la obra de Kummer. Su trabajo pone de manifiesto un hecho extraño de la investigación matemática, a saber, que es imposible predecir qué líneas de investigación conducirán a resultados útiles. Estudiando cuestiones subidamente teóricas, de la matemática más pura, llegó a formular conceptos de extraordinario valor y versatilidad en toda la matemática.

En particular, la teoría de Kummer ha permitido avanzar en el estudio del teorema de Fermat más que ningún otro trabajo anterior o posterior. Tan sólo unos cuantos años antes, las demostraciones de los casos n igual a 5 y n igual a 7 habían sido consideradas logros fundamentales; pero en 1847 Kummer pudo demostrar que el teorema era cierto para todos los exponentes primos menores que 37; y, por tanto, el teorema estaba probado para todo exponente menor que 37. Además, estuvo cerca de probar el teorema para todos los exponentes menores que 100.

Aunque muchos historiadores de la ciencia matemática hayan declarado que la teoría de Kummer surgió al estudiar éste el último teorema de Fermat, una lectura cuidadosa de la obra de Kummer y de su correspondencia muestra que su interés por el teorema de Fermat era bastante secundario. El principal objetivo de Kummer era encontrar solución a otro problema de la aritmética superior, llamada también teoría de números: el problema de las leyes de reciprocidad de grado superior, propuesto por Gauss. (El problema de las leyes de reciprocidad de grado superior consiste en generalizar para potencias mayores que dos la famosa ley de reciprocidad cuadrática, que Gauss demostró para segundas potencias. Enunciada brevemente, la ley de reciprocidad cuadrática establece que si p y q son enteros primos e impares, entonces las preguntas "¿Se diferencia p del cuadrado de un entero en un múltiplo de q?" y "¿Se diferencia q del cuadrado de un entero en un múltiplo de p?" guardan entre sí relación sencilla.) En 1847, los trabajos de Kummer sobre leyes de reciprocidad estaban todavía en sus primeras etapas; pero hacia 1859 consiguió un gran éxito, al demostrar un teorema general que fue culminación de su obra de teoría de números. El punto de vista tradicional, según el cual el interés de Kummer se centraba en el último teorema de Fermat, no anda, sin embargo, completamente errado, ya que el teorema de Fermat está emparentado de cerca con las leyes de reciprocidad. El propio Gauss, aunque negó siempre estar intrínsecamente interesado en el último teorema de Fermat, manifestó la esperanza de que apoyándose en sus resultados sobre leyes de reciprocidad de grado superior pudiera algún día deducir fácilmente el teorema.

Los números primos regulares

La teoría de Kummer, de 1847, era particularmente valiosa porque proporcionaba una condición suficiente para que un número primo impar fuese un exponente de validez del teorema de Fermat. Dicho de otra forma, si un número primo impar p cumple la condición de Kummer, entonces la ecuación $x^p + y^p = z^p$ carece de solución. Con terminología moderna, los números primos que verifican tal condición se llaman *regulares*. (Concretamente, un número primo p es regular si y solamente si no divide con solución exacta al numerador de ninguno de los primeros $p - 3$ números de la serie de fracciones hoy denominadas números de Bernouilli.) Si bien la condición de regularidad es suficiente para los exponentes del último teorema de Fermat, no es condición necesa-

ria. Hay otros números primos p que no son regulares y para los cuales se ha demostrado ser imposible que $x^p + y^p = z^p$. Todos los números primos menores que 100, excepto 37, 59 y 67 son regulares.

Kummer se precipitó al suponer que el conjunto de números primos regulares es infinito, pero pronto se dio cuenta de que no podía demostrar que así ocurre. De hecho, todos los esfuerzos posteriores para conseguir demostrarlo han fracasado, a pesar de que tanto la intuición como la evidencia experimental están a favor de la proposición de que existan infinitos números primos regulares. (Bastante extrañamente, se ha demostrado que hay infinitos número primos irregulares. Alrededor del 60 por ciento de los números primos alcanzables mediante las actuales técnicas de cálculo son regulares, y hay buenos motivos para creer que la mayoría de los números primos son regulares. Así pues, los números primos pueden dividirse en dos subconjuntos, los primos regulares y los primos irregulares, de los cuales, el que se supone sea el mayor de ambos conjuntos no se sabe demostrar que es infinito, mientras que sí se ha probado que su complementario es infinito.)

En años posteriores Kummer obtuvo nuevas condiciones suficientes para el último teorema de Fermat, que comprendían todavía más números primos, entre ellos, los primos 37, 59 y 67, que son irregulares. Desde los tiempos de Kummer hasta hoy se han encontrado condiciones suficientes todavía menos restrictivas, siendo la descubierta por el profesor H. S. Vandiver, de la Universidad de Texas, una de las más generales. Sin embargo, ni aún con las más amplias de tales condiciones ha podido demostrarse la validez del teorema para un conjunto infinito de números primos. Así pues, por improbable que pueda parecernos, es todavía concebible que el teorema de Fermat solamente sea verdadero para un conjunto finito de números primos, y, por consiguiente, que exista un número suficientemente grande M tal que el teorema sea falso para todos los números primos mayores que M.

Por otra parte, las condiciones suficientes actuales son ahora tan generales que incluyen a todos los números primos que han podido ensayarse. Dicho de otra forma, todos los numeros primos alcanzables mediante las actuales técnicas de cómputo han resultado ser exponentes de validez del teorema de Fermat. Los algoritmos precisos para verificar si un número primo cumple estas condiciones son muy sencillos, y a lo largo de los últimos 25 años se han realizado comprobaciones sistemáticas en grandes y modernos ordenadores, entre las que destacan las efectuadas por D. H. Lehmer, de la Universidad de California en Berkeley, R. G. Selfridge, de la Universidad de Florida, Wells Johnson, del Bowdoin College, y Samuel S. Wagstaff, de la Universidad de Illinois. A principios de 1970, los cálculos de Johnson establecían que el teorema de Fermat es verdadero para todos los exponentes primos menores que 30.000. Más recientemente, Wagstaff, trabajando con técnicas muy refinadas en un gran ordenador de la Universidad de Illinois, ha llevado el límite hasta más alla de 125.000.

La situación en el momento presente

Los cálculos anteriores muestran que si hay un contraejemplo al teorema de Fermat, es decir, si existen enteros x, y, z y p tales que $x^p + y^p = z^p$, dichos números serían tan enormes que sobrepasarían con mucho, no solamente la capacidad de cálculo manual, sino también la de los mayores sistemas de proceso de datos que puedan concebirse en un futuro remoto. Si p es un número primo situado más allá del límite de Wagstaff, pongamos por caso, del orden de 300.000, entonces puede demostrarse que es imposible la ecuación $x^p + y^p = z^p$ a menos que x, y o z sea divisible por p. Así pues, z^p debe ser mayor que $300.000^{300.000}$, número que consta, cuando menos, de un millón de dígitos. Otros resultados muestran que un contraejemplo requeriría números todavía más descabelladamente grandes.

Así que, en cierto sentido, el teorema de Fermat es empíricamente cierto. Si la ecuación $x^n + y^n = z^n$ tiene alguna solución, los números que comporta son tan inconcebiblemente grandes que los seres humanos nunca podran manejarlos. No obstante, desde el punto de vista matemático y filosófico, el tamaño de los números en nada afecta a la validez o falsedad del último teorema de Fermat. Cuando un matemático afirma que un resultado es válido para todos los números no se está refiriendo que el enunciado sea verdadero para los números que hayan podido encontrarse hasta entonces, o que puedan encontrarse algún día. Por el contrario, puesto que ni siquiera se ha podido demostrar que el teorema de Fermat sea verdadero para una colección infinita de exponentes primos, se podría decir que todo el trabajo realizado sobre el teorema no ha servido más que para comprobarlo en unos cuantos miles de casos particulares, y que no existe ningún indicio seguro de que el teorema sea verdadero.

¿En qué estado se encuentran actualmente los trabajos sobre el último teorema de Fermat? Aún siendo, sin duda alguna, uno de los más famosos problemas abiertos de la matemática contemporánea, no constituye objeto de grandes esfuerzos por parte de los investigadores, principalmente, porque no se sabe por dónde atacarlo. Ha salido indemne de los potentes métodos de Kummer, así como de sus numerosos perfeccionamientos posteriores; todavía no ha sido probado ni siquiera para una colección infinita de números primos. Si alguien pudiera proponer algún concepto nuevo que prometiera, con algún fundamento, progresos verdaderamente nuevos hacia la solución del problema, despertaría mucho interés y gran actividad. En el momento actual, sin embargo, el teorema de Fermat se encuentra aletargado.

Interés por el teorema

Durante los últimos 100 años, el teorema de Fermat ha sido uno de los temas favoritos de los matemáticos aficionados, juntamente con los problemas de la cuadratura del círculo y la trisección de un ángulo con regla y compás. La diferencia entre el teorema de Fermat y estos otros problemas es que los últimos se sabe ya que son imposibles, por lo que toda presunta solución puede ser rechazada de antemano. No hay, sin embargo, tal certeza en el teorema de Fermat. Por el contrario, existe una remota posibilidad de que Fermat descubriera una demostración elemental del teorema. Ello no obstante, raramente consiguen los aficionados demostrar el caso n igual a 3, como hizo Euler hace dos siglos, y mucho menos, los casos también conocidos n igual a 5, 7 u 11, que son mucho más difíciles.

Puede que la pertinacia de los matemáticos aficionados —muchos de ellos convencidos de que los resultados "están ahí", esperando que alguien se dé cuenta— guarde relación con la general ignorancia sobre la naturaleza del trabajo matemático. En este artículo he pretendido poner de manifiesto que las matemáticas, ni están agotadas, ni son meras rutinas; que, por el contrario, los matemáticos suelen ir a la deriva en un océano de preguntas cuyas respuestas desconocen. De hecho, al investigar en matemáticas, el razonamiento da tan tortuosos giros que el investigador suele terminar hallando respuesta a cuestiones muy diferentes de aquellas que se propuso elucidar. Además, como bien muestra la historia del teorema de Fermat, hasta los más grandes matemáticos cometen errores. El teorema de Fermat sigue siendo un intrigante misterio, y aunque su importancia histórica sea en gran medida accidental, continuará atrayendo la atención del mundo matemático en tanto no se descubra una demostración. Hay toda clase de razones para creer que, en el futuro, lo mismo que ocurrió en el pasado, los intentos de resolver el último teorema de Fermat serán origen de importantes progresos en matemáticas.

¿Desencadenó una supernova la formación del sistema solar?

Ciertos isótopos hallados en algunos meteoritos primitivos son probablemente restos de una enorme estrella que, al parecer, explotó cerca del sistema solar en formación un millón de años antes de que se originasen los meteoritos

David N. Schramm y Robert N. Clayton

Las explosiones estelares llamadas supernovas son sucesos raros: entre los cien mil millones de estrellas de nuestra galaxia se producen probablemente tres supernovas por siglo. Pese a su rareza, sin embargo, las supernovas representaron el papel de semilla en el origen del sistema solar. Mucha de la materia que se reunió para formar el Sol y los planetas era polvo y gas que había sido expulsado por supernovas a lo largo de un período de varios miles de millones de años. En la actualidad, hay indicios de una relación más íntima con una supernova particular. Parece que una enorme estrella explotó en la vecindad del embrionario sistema solar por la época en que el sistema se condensó.

Los indicios provienen del estudio de los elementos químicos en los meteoritos y, en particular, de medidas de la abundancia de los varios isótopos de ciertos elementos. En muestras de materia procedente de la clase más primitiva de meteoritos, la abundancia relativa de algunos isótopos es diferente de la que se encuentra en la Tierra, en rocas lunares y presumiblemente en todos los demás cuerpos del sistema solar. La anomalías referentes a isótopos se pueden explicar admitiendo que se inyectara en el sistema solar materia procedente de una supernova, sólo algunos millones de años antes de que los meteoritos se solidificasen. Hay poderosas razones para creer que la estrecha asociación entre la supernova y el nacimiento del sistema solar no fue una coincidencia. El colapso de una difusa nube de gas y polvo al formar el Sol y su sistema planetario pudo haber sido desencadenado por el cataclismo de la vecina explosión.

Los isótopos son clases de átomos que tienen el mismo número de protones (y, por tanto, representan el mismo elemento químico), pero diferente número de neutrones. El oxígeno, por ejemplo, tiene tres isótopos que son estables frente a la desintegración radiactiva. Todos ellos tienen ocho protones, pero tienen ocho, nueve o diez neutrones. Los isótopos se designan generalmente por el símbolo del elemento y el número de masa atómica, que no es más que la suma de los números de protones y neutrones. Así, los isótopos estables del oxígeno son el O-16, O-17 y O-18.

Se dice con bastante frecuencia que los isótopos de un elemento son químicamente idénticos. Y es verdad que los procesos químicos no pueden transformar un isótopo en otro, de la misma manera que no pueden transformar un elemento en otro. En la mayoría de los casos es incluso una aproximación razonable decir que los procesos químicos no distinguen entre isótopos, de modo que cualquier isótopo entra en la misma reacción como lo hacen los otros. En la formación de los minerales, por ejemplo, los diferentes elementos se separan de acuerdo con sus propiedades químicas, pero todos los isótopos de un elemento dado permanecen ordinariamente juntos. De hecho, los isótopos son tan semejantes a este respecto que el separarlos constituye un difícil y costoso proceso.

La afirmación de que los isótopos son químicamente equivalentes entre sí es, sin embargo, una aproximación: la mayoría de los isótopos muestran pequeñas diferencias en sus propiedades químicas. Todas las diferencias están estrictamente determinadas por la masa atómica, que tiene una ligera influencia sobre la velocidad y el punto de equilibrio de las reacciones químicas. El efecto rara vez es grande, pero puede ser importante. Por ejemplo, el punto hasta el cual los isótopos más pesados del oxígeno se incorporan a los sedimentos marinos es sensible a la temperatura, y así la abundancia de isótopos en tales sedimentos se puede usar para determinar la temperatura de los antiguos océanos.

Por depender siempre la separación química de los isótopos de la masa atómica, aquélla sigue una norma sencilla. Si la relación de O-17 a O-16 aumenta ligeramente como consecuencia de un proceso químico, la relación de O-18 a O-16 debe aumentar en doble cantidad, ya que la diferencia de masa es el doble. Conociendo esta relación, es posible compensar los efectos de la separación química y recuperar la "verdadera" abundancia isotópica.

Cuando se toman en cuenta esos pequeños efectos químicos, se halla que las abundancias relativas de isótopos se mantienen constantes con gran aproximación. El oxígeno terrestre es 99,756 por cien de O-16, 0,039 por cien de O-17 y 0,205 por cien de O-18. Esas mismas proporciones

GRANOS MINERALES de un meteorito; comprenden elementos cuya composición isotópica difiere de la del resto de la materia conocida en el sistema solar. Las muestras de la página opuesta proceden del meteorito de Allende, que cayó en México en 1969. En la microfotografía superior se ha irradiado un corte pulido de roca con electrones, de modo que se puedan identificar los minerales por el color característico de su luminiscencia. Un mineral de particular interés es la anortita, que es desusadamente rico en aluminio y pobre en magnesio. Los granos de anortita aparecen de color azul claro; otros minerales son la melilita (azul obscuro o púrpura), piroxeno (rojo) y espinela (amarillo). En la parte de abajo se ven tres cristales de anortita en una microfotografía hecha con filtros polarizadores cruzados, que prestan su falso color a los cristales. Cerca de la parte superior izquierda de un cristal hay un pequeño cráter hecho por un haz de iones que se empleó para extraer elementos para su análisis isotópico. El análisis de esta muestra reveló un exceso del isótopo de magnesio con número de masa atómica 26, creado por la desintegración radiactiva del aluminio 26. El isótopo de aluminio debió haberse producido en una explosión de supernova poco tiempo antes de que el meteorito se condensara. (Las explosiones estelares llamadas supernovas son sucesos raros, entre los cien mil millones de estrellas de nuestra galaxia se producen probablemente tres supernovas por siglo.) Las fotografías fueron realizadas por Ian M. Steele *(arriba)* e Ian D. Hutcheon *(abajo)*, ambos de Chicago.

se observan (tras corregir los efectos químicos) en oxígeno de la atmósfera, del agua del mar y de rocas sedimentarias o ígneas. La cantidad de oxígeno en esas substancias, por supuesto, varía ampliamente, pero el oxígeno en sí parece ser idéntico en composición isotópica.

Las relaciones de isótopos para el oxígeno y otros elementos se han medido en rocas lunares y se han hallado ser esencialmente las mismas que se encuentran en la Tierra. De hecho, la teoría más aceptada sobre el origen del sistema solar haría esperar que se observasen las mismas relaciones si se pudiera examinar materia procedente de cualquiera de los otros planetas. En esta teoría, el sistema solar se concibe como emergiendo del colapso gravitatorio de una nube de gas y polvo, difusa y turbulenta. Los elementos más ligeros y más abundantes en la nube, el hidrógeno y el helio, fueron los materiales primordiales, pero los elementos más pesados eran, sobre todo, residuos de supernovas.

La materia expulsada por todas las supernovas que contribuyeron a la nube protosolar no tiene necesariamente que haber sido idéntica en composición isotópica o incluso en composición elemental: de hecho es probable que cada supernova emitiese una mezcla característica de isótopos y elementos. Sin embargo, una vez que esas diversas contribuciones se reunieron en la nube quedaron completamente uniformadas. Hacia la época en que el sistema solar se condensó a partir de la nube, la composición isotópica era homogénea y reflejaba tan sólo el contenido isotópico medio de todos los residuos de supernovas. El reciente descubrimiento de materia con una composición isotópica que difiere de la media del sistema solar indica que al menos una supernova explotó más tarde que el resto, demasiado

NUCLEO ESTELAR

SUPERNOVA

$^{12}C \rightarrow {}^{12}C + {}^{4}He \rightarrow {}^{16}O$

$^{4}He + {}^{4}He \rightarrow {}^{8}Be$; $^{8}Be + {}^{4}He \rightarrow {}^{12}C$; $^{12}C + {}^{4}He \rightarrow {}^{16}O$

$^{16}O + P \rightarrow {}^{17}F \rightarrow {}^{17}O + \beta$

$^{17}O + P \rightarrow {}^{4}He + {}^{14}N$

$^{17}O + P \rightarrow {}^{18}F \rightarrow {}^{18}O + \beta$

$^{18}O + {}^{4}He \rightarrow {}^{22}Ne$; $^{22}Ne + {}^{4}He \rightarrow {}^{25}Mg + n$

$^{4}He + {}^{4}He \rightarrow {}^{8}Be$; $^{8}Be + {}^{4}He \rightarrow {}^{12}C$; $^{12}C + {}^{12}C \rightarrow {}^{20}Ne + {}^{4}He \rightarrow {}^{24}Mg$

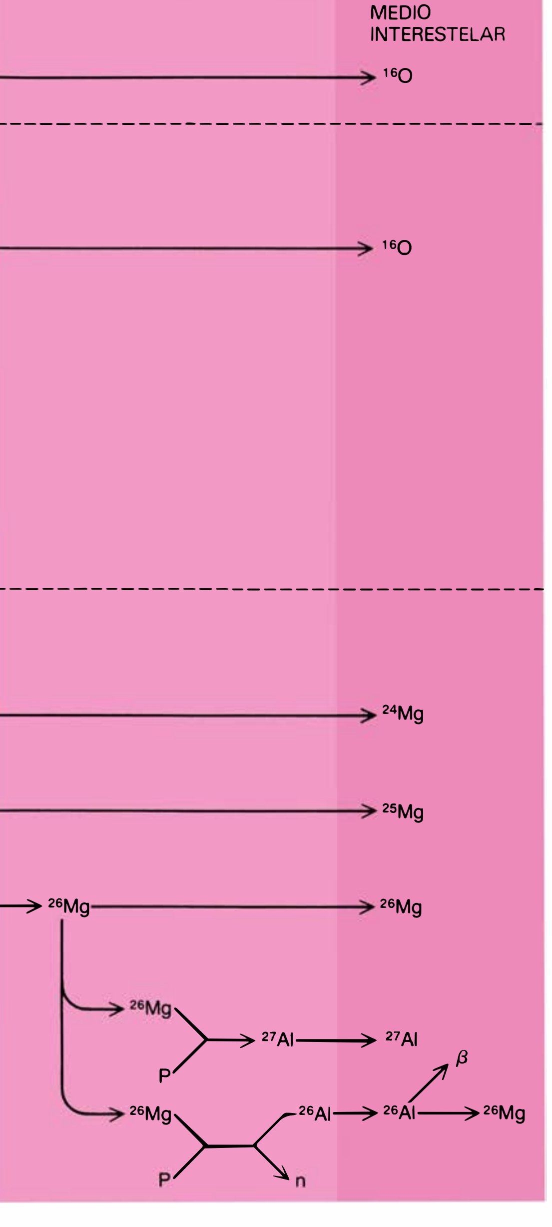
$^{24}Mg + n \rightarrow {}^{25}Mg$

$^{25}Mg + n$

LA SINTESIS DE ELEMENTOS en los que se han encontrado anomalías isotópicas está relacionada con reacciones nucleares, bien en el núcleo de una estrella de gran masa o en la explosión de tal tipo de estrella. En cualquiera de los casos se necesita una supernova que expulse el material. Los núcleos inestables pueden transformarse más tarde por desintegración espontánea después de haber sido expulsados. Se muestran algunas secuencias de reacciones representativas para dos elementos, el oxígeno y el magnesio. Los isótopos más importantes de ambos elementos, O-16 y Mg-24, ocupan un importante lugar en una vía fundamental de la nucleosíntesis; eso explica, en parte, su abundancia. Los otros isótopos se producen de maneras más marginales. El oxígeno 17 y el 18 se forman en las estrellas durante un ciclo de reacciones cuyo producto final es el helio 4. Los núcleos de O-17 y O-18, sin embargo, no pueden sobrevivir en las altas temperaturas reinantes en una supernova. El O-17 es un núcleo frágil que se puede desintegrar de varias maneras; el O-18 absorbe fácilmente un núcleo de helio para transformarse en neón 22. Algo de O-17 y O-18 se pueden escapar de la estrella en el viento estelar (*no representado*), pero el oxígeno expulsado por una supernova es esencialmente O-16 puro. El Mg-25 y Mg-26 se pueden formar durante una supernova como resultado de captura de neutrones por parte del abundante Mg-24. Algunos de estos isótopos pueden dar lugar al Al-27, único isótopo estable del aluminio. Otras reacciones, tales como la

tarde para que la materia que expulsó se mezclara completamente con todas las contribuciones previas.

Aunque las relaciones de isótopos no quedan alteradas de un modo substancial por las reacciones químicas, se puede esperar que hoy día persistan relaciones anómalas sólo en materia que ha sufrido pocas modificaciones químicas. La razón es que la mezcla de materiales de diversas fuentes que tuvo lugar en la nube protosolar ha continuado operando en los sistemas químicos de los planetas más activos. Supóngase, por ejemplo, que en la región de la nube donde se formó la Tierra, un uno por cien del oxígeno presentaba una composición isotópica anómala. Habría sido posible encontrar algún indicio de ese oxígeno particular recién condensado el planeta, hace 4600 millones de años, pero hoy no podría detectarse. El oxígeno en la atmósfera, en el agua, en las rocas y en los organismos vivientes sufre de manera continua un trama interconectada de ciclos químicos. Precisamente porque los isótopos se semejan en su composición química, el oxígeno isotópicamente desusado se mezclaría con todo el resto. Las anomalías no se podrían reconocer hoy sólo mediante análisis de materia terrestre, pero se podrían observar comparando oxígeno de la Tierra con oxígeno de otro planeta.



absorción simultánea de un protón y emisión de un neutrón, producen pequeñas cantidades del isótopo radiactivo Al-26. El Al-26 se forma en su mayor parte en las supernovas, y se desintegra con una semivida de 720.000 años dando Mg-26. El exceso de Mg-26 con Al-27 en un meteorito se puede interpretar como prueba de que el meteorito se condensó no más tarde de un millón de años tras la explosión de la supernova; si el tiempo hubiera sido mayor, casi todo el Al-26 se habría desintegrado.

Un lugar adecuado para buscar relaciones isotópicas desusadas serían los cometas, que quizá han sido muy poco alterados por procesos químicos. Los cometas, sin embargo, resultan inaccesibles en absoluto. Las mejores fuentes de materia del sistema solar primitivo fácilmente disponibles son los meteoritos, que al parecer nunca formaron aglomeraciones lo bastante grandes para que llegaran a establecerse en ellos ciclos geoquímicos. Entre los meteoritos, la clase más primitiva la constituyen los llamados condritas carbonáceas. Se distinguen por la presencia de carbono y de pequeñas inclusiones redondas llamadas cóndrulos, que presentan indicaciones de haber estado fundidos en alguna época. Se cree que las condritas carbonáceas se solidificaron en los albores de la historia del sistema solar y que muy poco ha sucedido después que pudiera alterarlas.

La mayor parte de las anomalías isotópicas recién descubiertas se encontraron por primera vez en materia procedente de una única condrita carbonácea, que cayó en 1969 cerca de la localidad de Pueblito de Allende, en el Norte de México. Se conoce ahora por el "meteorito de Allende". En otras condritas carbonáceas se han encontrado las mismas relaciones características, de modo que éstas no son exclusivas del material de Allende; por el contrario, son probablemente comunes a todos los meteoritos de esta clase. Una interesante pregunta es por qué los primeros exámenes de material meteorítico fueron incapaces de descubrir discrepancias isotópicas. Una parte de la explicación es que el meteorito de Allende era mucho mayor que ninguna de las otras condritas carbonáceas, de modo que había más materia disponible para su estudio. El meteorito de Allende se rompió en muchos fragmentos, la mayoría de los cuales eran del tamaño del puño, pero el cuerpo original se estima haber tenido una masa de más de dos toneladas. Otro factor es que el meteorito de Allende cayó en un momento en que varios laboratorios se estaban equipando para analizar materia lunar traída por las expediciones Apolo. En esos laboratorios, las relaciones isotópicas en el material de Allende se podían medir con mayor precisión de la que había sido posible con meteoritos anteriores. Por otra parte, las anomalías podrían haber sido halladas varios años antes si tales laboratorios no hubieran estado tan ocupados analizando las rocas lunares.

La primera anomalía descubierta en el material de Allende correspondía a los isótopos de oxígeno. Fue encontrada por uno de nosotros (Clayton) en colaboración con Toshiko Mayeda y Lawrence Grossman, de la Universidad de Chicago. El oxígeno es un elemento corriente (es un componente de la mayoría de los minerales), y las relaciones de isótopos de oxígeno se han venido determinando corrientemente en una gran variedad de materiales del sistema solar. Todas las medidas anteriores concordaban con la hipótesis de una nube protosolar que fuera homogénea en su totalidad. El descubrimiento de una relación diferente en algunas muestras del meteorito de Allende fue sorprendente.

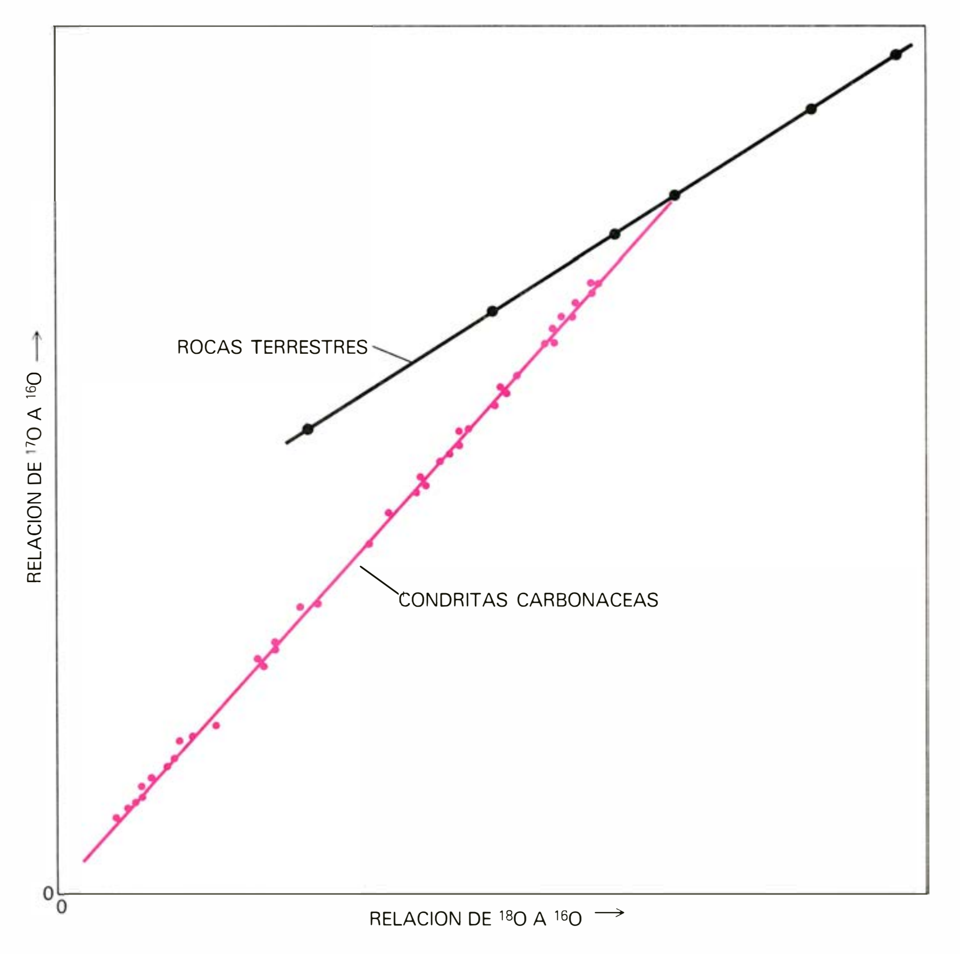
Las relaciones isotópicas se midieron con un espectrómetro de masas, instrumento que separa los átomos de acuerdo con sus masas. El oxígeno del material de Allende se aisló primero por métodos químicos, y los átomos de oxígeno se ionizaron de tal manera que pudieran ser acelerados. Los iones acelerados pasaron entonces a través de un campo magnético, que los desvió en una cuantía deteminada por su masa. Colocando adecuadamente un detector de iones más allá del imán, se podía contar el número de átomos de cada isótopo.

Como se indicó arriba, el oxígeno terrestre contiene más del 99 por cien de O-16; los isótopos más pesados constituyen trazas. En el material de Allende se halló que la proporción de O-16 era todavía más alta. La comparación de las varias relaciones en cuestión demostró que la discrepancia no podía deberse a fraccionamiento químico: no era proporcional a la diferencia de masas. En efecto, la relación de O-17 a O-18 permaneció imperturbada. El oxígeno del meteorito parecía ser una mezcla de dos componentes, uno que tenía la composición isotópica del oxígeno normal en el sistema solar y otro compuesto de O-16 puro. En algunas muestras, el O-16 puro representaba hasta el cinco por ciento del oxígeno.

Puesto que el O-16 es ya el más abundante isótopo del oxígeno, la adición de un cinco por cien más de él cambia la relación isotópica de forma muy ligera: de 99.756 por cien de O-16 a 99.768 por cien. La relación se mantiene lo suficiente-

mente constante en otros materiales, sin embargo, para que la pequeña discrepancia se deba considerar representativa. Implica que la nube protosolar no era enteramente homogénea y que algunos de sus componentes tenían diferente composición isotópica. Si el O-16 "extra" hubiera estado en forma gaseosa, se habría mezclado con el resto de la nube y se habría diluido muy rápidamente. Es más probable que el oxígeno anómalo entrase a formar parte de la nube en forma de combinación química dentro de granos sólidos.

Cabe proponer dos orígenes posibles de tales granos. Podrían haber sido granos primitivos, y en ese caso todos los granos interestelares podrían haber tenido una composición isotópica diferente de la del gas interestelar. Esa diferencia no se detectaría en otros cuerpos del sistema solar, porque los elementos procedentes de la fase sólida y de la gaseosa se han mezclado íntimamente. O bien podría haber sucedido que la mayoría de los granos y el gas en la nube hubieran tenido idéntica composición isotópica, con sólo unos pocos granos anómalos, que habrían sido añadidos demasiado tarde para mezclarse por completo. Estos granos se podrían haber conservado en ciertos meteoritos mientras que en la Tierra y otros planetas se habrían descompuesto y recombinado químicamente. Como se verá más abajo, existen poderosas indicaciones que favorecen la última hipótesis, pero los isótopos del oxígeno por sí solos no suministran una base suficiente para escoger entre las dos posibilidades. Debe subrayarse el hecho de que los isótopos de oxígeno no son portadores de información sobre la fecha en que se formaron los granos peculiares. Los tres isótopos de oxígeno son estables, y, por tanto, se pudo haber "congelado" una relación isotópica desusada (tal como 100 por cien de O-16) en algunos granos de polvo cientos de millones de años antes de que tales granos se incorporasen al sistema solar. El único requisito para explicar las observaciones es que los granos no hayan tenido posibilidad de mezclarse completamente con materiales más típicos del sistema solar.



LOS ISOTOPOS DE OXIGENO en muchas muestras de material meteorítico tienen una composición notablemente diferente de la que presentan en rocas terrestres. Los procesos químicos pueden alterar, de una forma ligera, todas las relaciones isotópicas, pero la variación es siempre proporcional a la masa atómica de los isótopos. Por esta razón, la composición isotópica del oxígeno terrestre no es completamente uniforme, si bien las variaciones siguen una regla sencilla. Si la relación de O-17 a O-16 se incrementa en una unidad, la relación de O-18 a O-16 debe aumentar entonces en dos unidades, porque la diferencia de masa es dos veces mayor. Así, una gráfica de las dos relaciones es una recta (*en negro*) con pendiente igual a 1/2. En los meteoritos llamados condritas carbonáceas, incluyendo el de Allende, se ha descubierto una relación diferente. En estas muestras, si la relación de O-17 a O-16 aumenta en una unidad, también lo hace la de O-18 a O-16; en otras palabras, la recta tiene pendiente igual a 1 (*en color*). Este descubrimiento sugiere que el oxígeno en las condritas carbonáceas es una mezcla de dos componentes. Uno de ellos es el oxígeno terrestre normal. El otro es O-16 puro, que se ha añadido en varias proporciones hasta aproximadamente un 5 por ciento. La fuente más razonable para el O-16 extra es una supernova que explotó demasiado tarde para que sus residuos se mezclasen completamente con el resto de la nube protosolar. (Ilustración de Kelvin.)

Se puede dar la fecha de otra anomalía isotópica observada por primera vez en el meteorito de Allende. Está constituida por pruebas de que, cuando el meteorito se formó, incluía cierta cantidad de un isótopo radiactivo del aluminio, Al-26. La prueba se halló en experiencias realizadas por Typhoon Lee, entonces en el Instituto de Tecnología de California y ahora en la Universidad de Chicago, y por Dimitri A. Papanastassiou y Gerald J. Wasserburg de dicho Instituto. El aluminio tiene sólo un isótopo estable, con 13 protones y 14 neutrones y, por tanto, un número másico de 27. El aluminio terrestre está constituido enteramente por este isótopo. El aluminio 26, con 13 protones y 13 neutrones, tiene una semivida de 720.000 años, lo que significa que después de tal período de tiempo se habrán desintegrado la mitad de los átomos de la muestra de Al-26. La desintegración tiene lugar mediante emisión beta positiva: se emiten un positrón (o antielectrón) y un neutrino, y un protón del núcleo se transforma en neutrón. En esta transformación el número másico del átomo no cambia, pero el nuevo núcleo tiene solamente 12 protones y 14 neutrones; ello lo convierte en un núcleo de magnesio 26, que es un isótopo estable.

Supongamos que cierta cantidad de Al-26 fuera introducida en la nube protosolar inmediatamente antes de que se condensaran las condritas carbonáceas. Tendría esencialmente las mismas propiedades químicas que el Al-27 y, por tanto, aparecería en relación constante con el Al-27 en todos los minerales que incluyen aluminio. Tras unos pocos millones de años, sin embargo, la mayor parte del Al-26 se habría desintegrado, y tras los 4600 millones de años que han transcurrido desde la formación del sistema solar, todo él se habría convertido en Mg-26. Si la roca nunca estuvo fundida o sometida a otros procesos que pudieran separar los elementos, la cantidad de Al-26 presente en un principio se podría medir por la simple determinación de la cantidad de Mg-26.

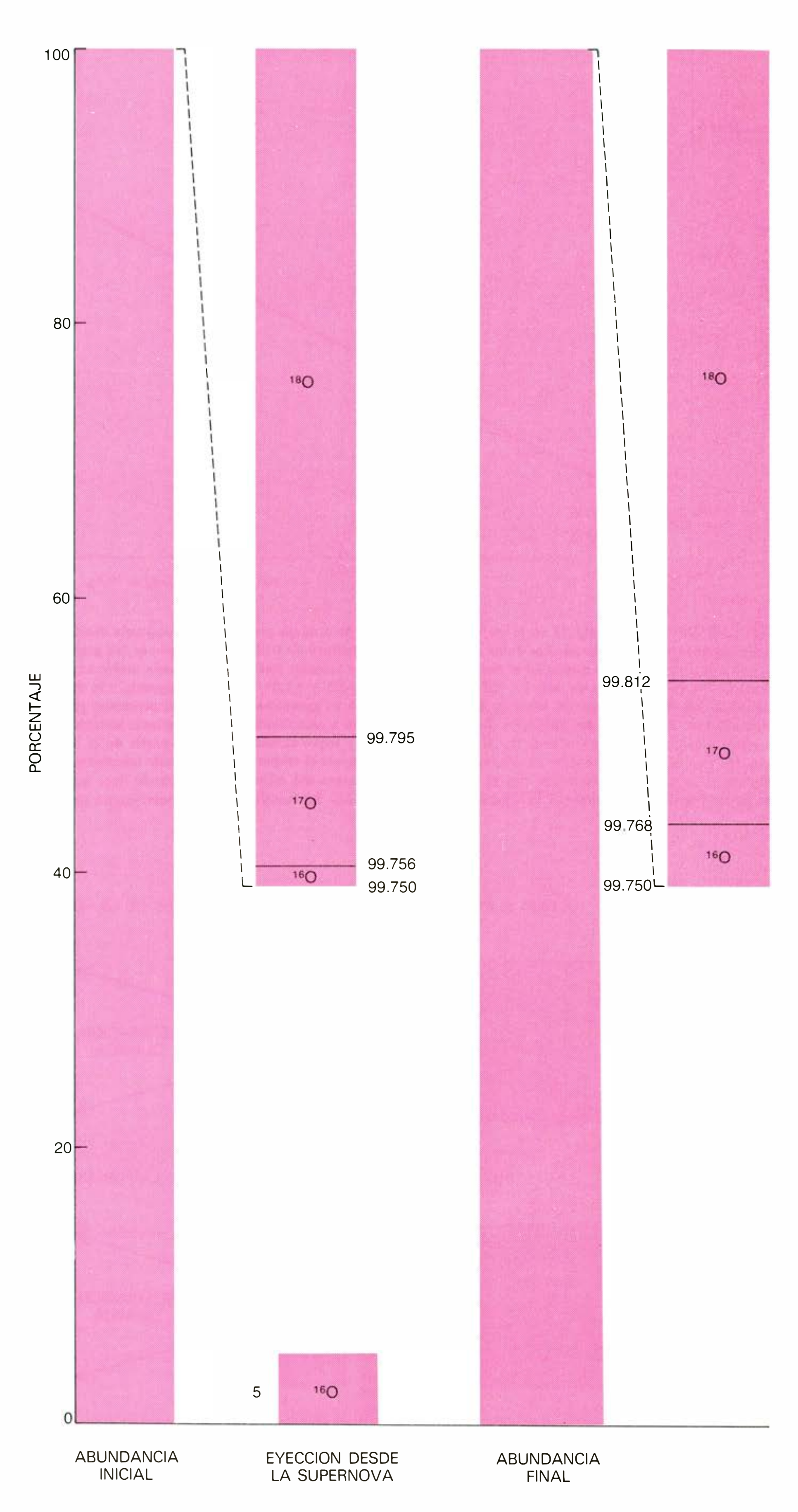
En la práctica, la medida no es tan sencilla porque virtualmente todos los minerales incluyen magnesio procedente de otras fuentes; de hecho, el magnesio abunda mucho más que el aluminio. Hay

tres isótopos estables de magnesio con números másicos 24, 25 y 26; la composición isotópica normal es 78,99 por cien de Mg-24, 10 por cien de Mg-25 y 11,01 por cien de Mg-26. Por tanto, cualquier Mg-26 creado por desintegración de Al-26 se podrá detectar sólo como exceso sobre la abundancia normal de tal isótopo.

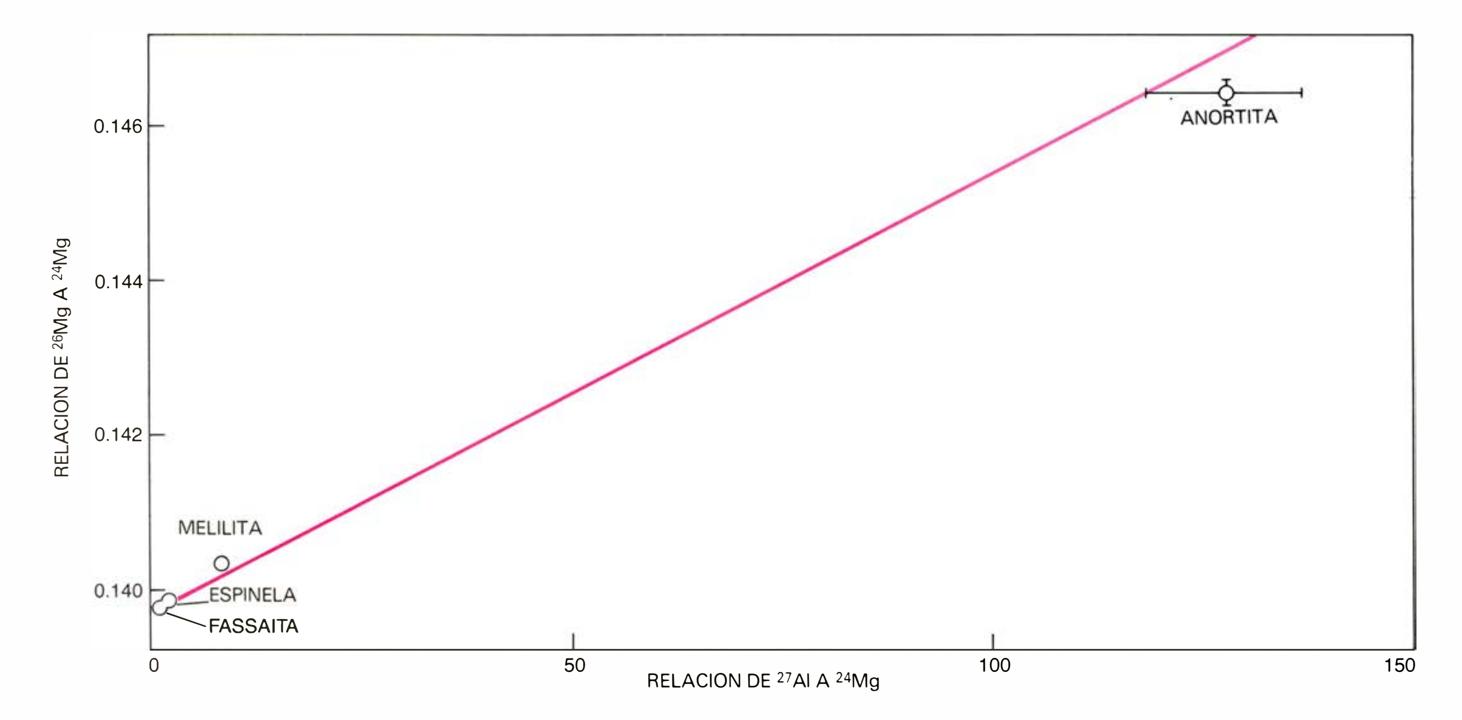
Esta clase exacta de incrementos fue observada por Lee, Papanastassiou y Wasserburg en el material de Allende. En algunas muestras, el porcentaje de Mg-26 aumentaba con respecto a su nivel normal desde 11,01 a unos 11,5. Sólo se observaron tan grandes anomalías en minerales tales como la anortita, que es rica en aluminio y pobre en magnesio; la mayor parte de las anomalías eran mucho menores.

El incremento de Mg-26 no pudo deberse a fraccionamiento químico; en tal caso habría habido un aumento similar, aunque más pequeño, en Mg-25, que no se observó. Quedaba por demostrar, sin embargo, que el exceso de Mg-26 procedía de la desintegración de Al-26; varias otras reacciones nucleares podrían haber dado origen al mismo producto. Sin embargo, se pudo establecer el origen del Mg-26 comparando el exceso de éste con la relación elemental de aluminio a magnesio en cada grano mineral. Tales medidas parecen mostrar una sencilla relación lineal: cuanto más grande la relación Al/Mg, tanto mayor el exceso de Mg-26. Si el Mg-26 hubiera tenido otra procedencia, no habría motivo para esperar correlación alguna. La correlación del contenido de aluminio y exceso de Mg-26 en muchos minerales hace incluso posible estimar la cantidad de Al-26 presente en un principio. Parece que el isótopo Al-26 era un contaminante escaso: aproximadamente un átomo por cada 20.000 átomos del isótopo estable Al-27.

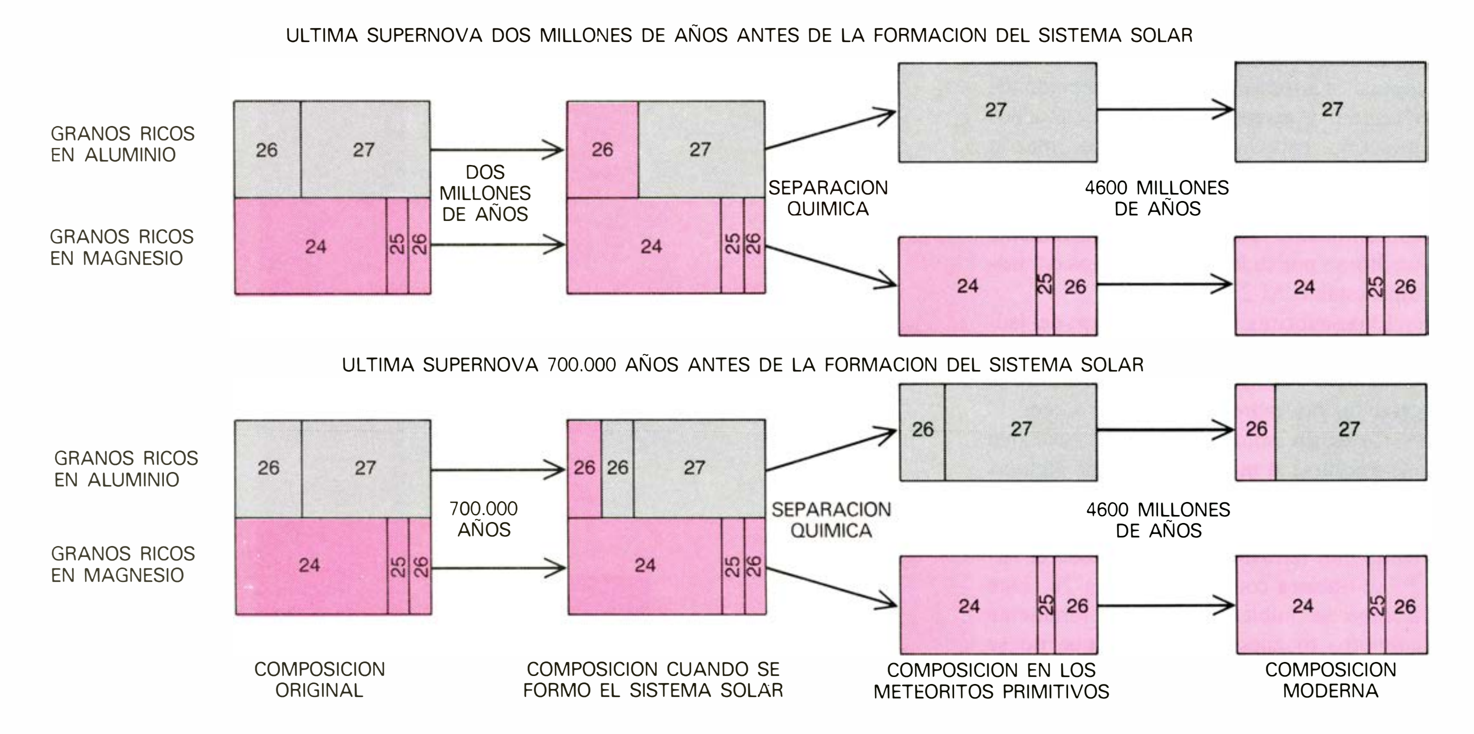
El descubrimiento de un exceso del isótopo Mg-26 resultó de particular relevancia porque suministraba una fecha para la creación del material anómalo. Como el Al-26 es un isótopo radiactivo, tuvo que incorporarse al mineral a lo sumo algunas semividas, o unos pocos millones de años, después que aquél se formara. Si hubiera habido un retraso algo mayor, todo el Al-26 se hubiera convertido en Mg-26. Este isótopo se hubiera mezclado libremente con otro magnesio, de modo que no se hubiera detectado correlación entre aluminio y magnesio. La conclusión más razonable es que el Al-26 se creó en una supernova no más de unos pocos millones de años antes de que se formase el sistema solar. Se estima que la condensación de una estrella del tamaño del Sol tarda unos diez millones de años, de modo que, en la escala de tiempo de la formación de estre-



EL EXCESO DE OXIGENO 16 sólo altera ligeramente la composición isotópica del elemento porque el oxígeno normal ya contiene más del 99 por ciento de O-16. La composición de oxígeno terrestre, una vez corregida de fraccionamiento químico, se da en la izquierda, con el 0,25 por ciento más alto representado en una escala mayor. La mezcla de tal gas con un 5 por ciento de oxígeno 16 puro eleva la proporción de oxígeno 16 sólo de 99.756 por ciento a 99.768 por cien. Aunque el aumento es numéricamente pequeño, resulta importante porque la escala de otras variaciones en el seno de las relaciones isotópicas es menor.



EXCESO DE MAGNESIO 26 en el meteorito de Allende; se produjo por la desintegración del aluminio 26. Los datos corresponden a minerales identificados en una única inclusión rocosa en el meteorito. El magnesio terrestre normal contiene 78,99 por ciento de Mg-24, 10 por ciento de Mg-25 y 11,01 por ciento de Mg-26; por tanto, la relación de Mg-26 a Mg-24 es aproximadamente 0,139. Algunos de los minerales de Allende se acercan a esta relación, particularmente la fassaíta y la espinela. En la melilita hay un ligero exceso de Mg-26, y en la anortita el exceso es substancial. Lo que demuestra el origen del Mg-26 es la estrecha correlación con el contenido de aluminio del mineral (expresado cuantitativamente por la relación de Al-26 a Mg-24). La fassaíta y la espinela contienen ambas aluminio y magnesio, pero la abundancia de magnesio es tan grande que el pequeño exceso de Mg-26 procedente de Al-26 no se puede medir experimentalmente. La melilita es rica en aluminio y pobre en magnesio y la anortita es rica en aluminio y casi carece de magnesio, de modo que presentan grandes aumentos de la relación de Mg-26 a Mg-24. El enriquecimiento isotópico es proporcional a la relación de abundancia de los elementos. A partir de la constante de proporcionalidad o de la pendiente de la gráfica, puede calcularse la abundancia original de Al-26: el aluminio en la inclusión de Allende tiene un átomo de Al-26 por cada 20.000 átomos de Al-27. Las mediciones fueron tomadas por Lee, Papanastassiou y Wasserburg, del Cal. Tech.



DESINTEGRACION DE ALUMINIO 26; proporciona una manera de establecer la fecha de la supernova que produjo el isótopo. Como el Al-26 tiene una semivida de sólo 720.000 años poco de él sobrevive más allá de unos dos millones de años. Si la última supernova que contribuyó al sistema solar se hubiera producido más de dos millones de años antes de que se condensara el sistema solar, casi todo el Al-26 se hubiera desintegrado antes de que pudiese quedar encerrado en cuerpos sólidos. Las reacciones químicas hubieran separado el aluminio y el magnesio, y no se habrían podido detectar anomalías isotópicas. Si la última supernova que contribuyó se hubiera producido algo más tarde, cuando los meteoritos se solidificaron habrían conservado una fracción considerable del isótopo radiactivo. En la separación química de los elementos durante la condensación, esa fracción hubiera permanecido con el resto del aluminio, y cuando finalmente se desintegrase en el meteorito, habría sufrido pocas perturbaciones químicas. Así, sólo si la supernova se hubiera producido suficientemente tarde, hubiera creado un exceso de Mg-26 en las rocas ricas en aluminio. La abundancia de Al-26 se ha exagerado para más claridad.

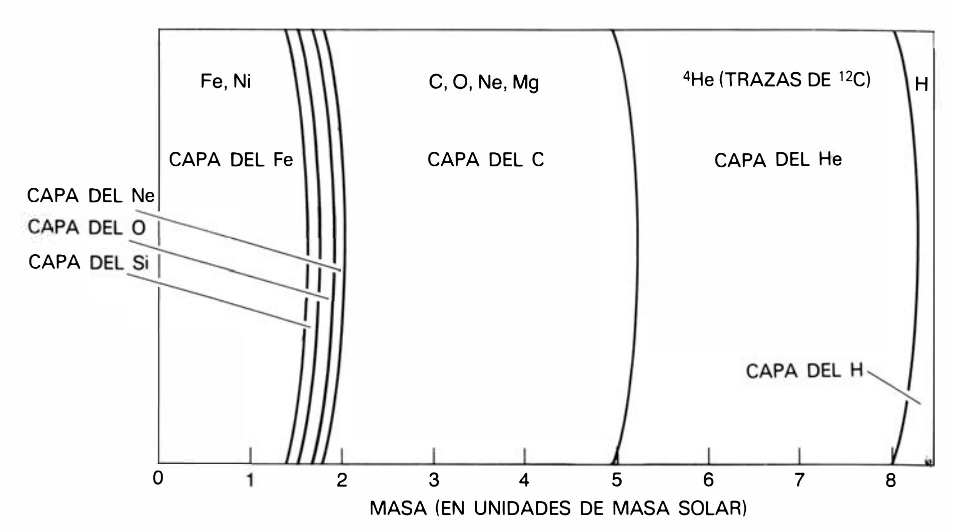
llas, los dos sucesos fueron esencialmente simultáneos.

Cabe imaginar otras explicaciones, pero son incapaces de justificar todos los datos. Por ejemplo, el Al-26 podría haber sido expulsado de una supernova muy anterior y haberse condensado inmediatamente con otros elementos formando granos sólidos, que más tarde se incorporaron sin alteración a los meteoritos. La radiactividad del Al-26 se habría extinguido entonces antes de la formación del sistema solar, pero la relación macroscópica entre aluminio y magnesio no se habría conservado. La anomalía del Al-26 se mantiene uniformemente en volúmenes de material de tamaño centimétrico y es improbable que los granos primitivos pudieran haber sido tan grandes; además, otros elementos observados en los granos presentan relaciones isotópicas que concuerdan con el promedio del sistema solar. Una supernova no hubiera producido Al-26 y, a la vez, los otros elementos con abundancias que imitan con exactitud la composición de la nube protosolar.

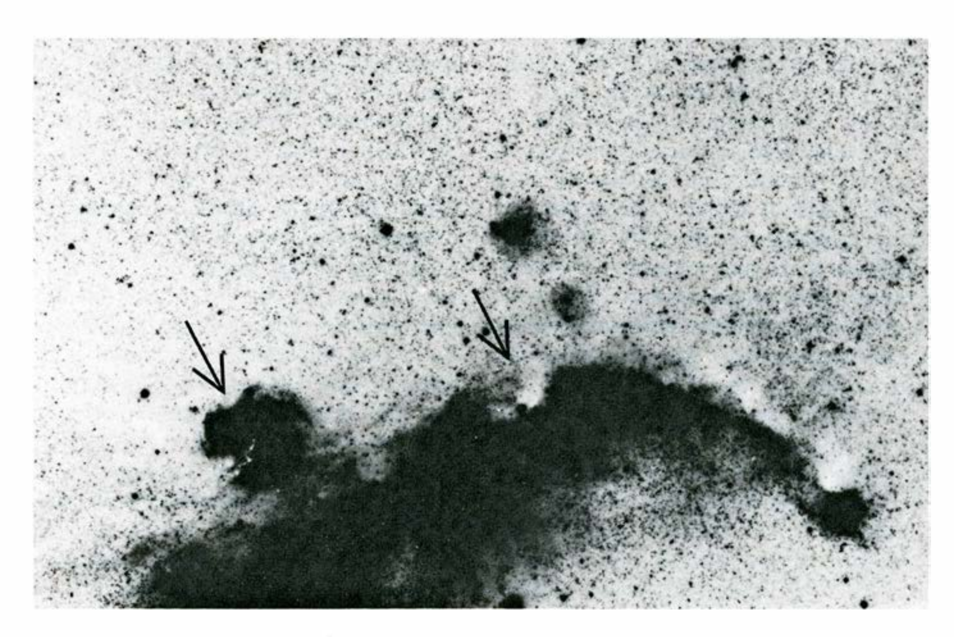
Otra posibilidad es que el Al-26 se produjese dentro del intervalo de tiempo considerado, de varios millones de años, antes de la solidificación, pero no en una supernova. Podría haberse producido dentro de la propia nube protosolar si alguno de los elementos ya presentes hubiera estado sometido a una radiación intensa. Pero no se conoce una tal fuente de radiación, y si hubiera existido, habría creado muchos otros tipos de anomalías isotópicas que no se han observado.

Para comprender por qué una supernova es la fuente más probable del Al-26 (y de muchos otros isótopos) es necesario considerar la manera como los elementos químicos se formaron. La "gran explosión" con la que comenzó el universo hace quince a veinte mil millones de años sólo produjo muy poca nucleosíntesis: la materia creada en aquel suceso inicial era casi exclusivamente hidrógeno y helio y quizás indicios de otros pocos elementos ligeros, tales como el litio. Todos los elementos más pesados en el universo se fueron elaborando a lo largo de diversas fases del ciclo de vida de las estrellas. Una de las fases iniciales en la nucleosíntesis estelar es la fusión de los átomos de hidrógeno para formar helio. Esto puede efectuarse mediante una reacción directa entre núcleos de hidrógeno y también mediante un ciclo más complicado en el que varios isótopos de carbono, nitrógeno y oxígeno sirven de catalizadores efectivos sobre los cuales se edifican los átomos de helio. En una fase posterior que requiere temperaturas más altas, los átomos de helio se fusionan para construir elementos más pesados, los cuales se fusionan, a su vez, para formar núcleos todavía más pesados. Tal fusión genera en forma particularmente abundante aquellos núcleos que pueden ser considerados múltiplos del isótopo He-4. El primero de ellos, el berilio 8, es inestable y se desintegra si no absorbe otro núcleo de helio, pero el resto son todos isótopos abundantes. Entre ellos se cuentan el carbono 12, oxígeno 16, neón 20, magnesio 24 y silicio 28. Si hay temperaturas todavía más altas, pueden provocarse otras reacciones de fusión. En el proceso de más alta temperatura que actúa en las estrellas normales, se arrancan núcleos de helio de algunos núcleos de silicio y se añaden a otros núcleos de silicio. De esta manera puede con el tiempo llegar a crearse hierro, de 26 protones.

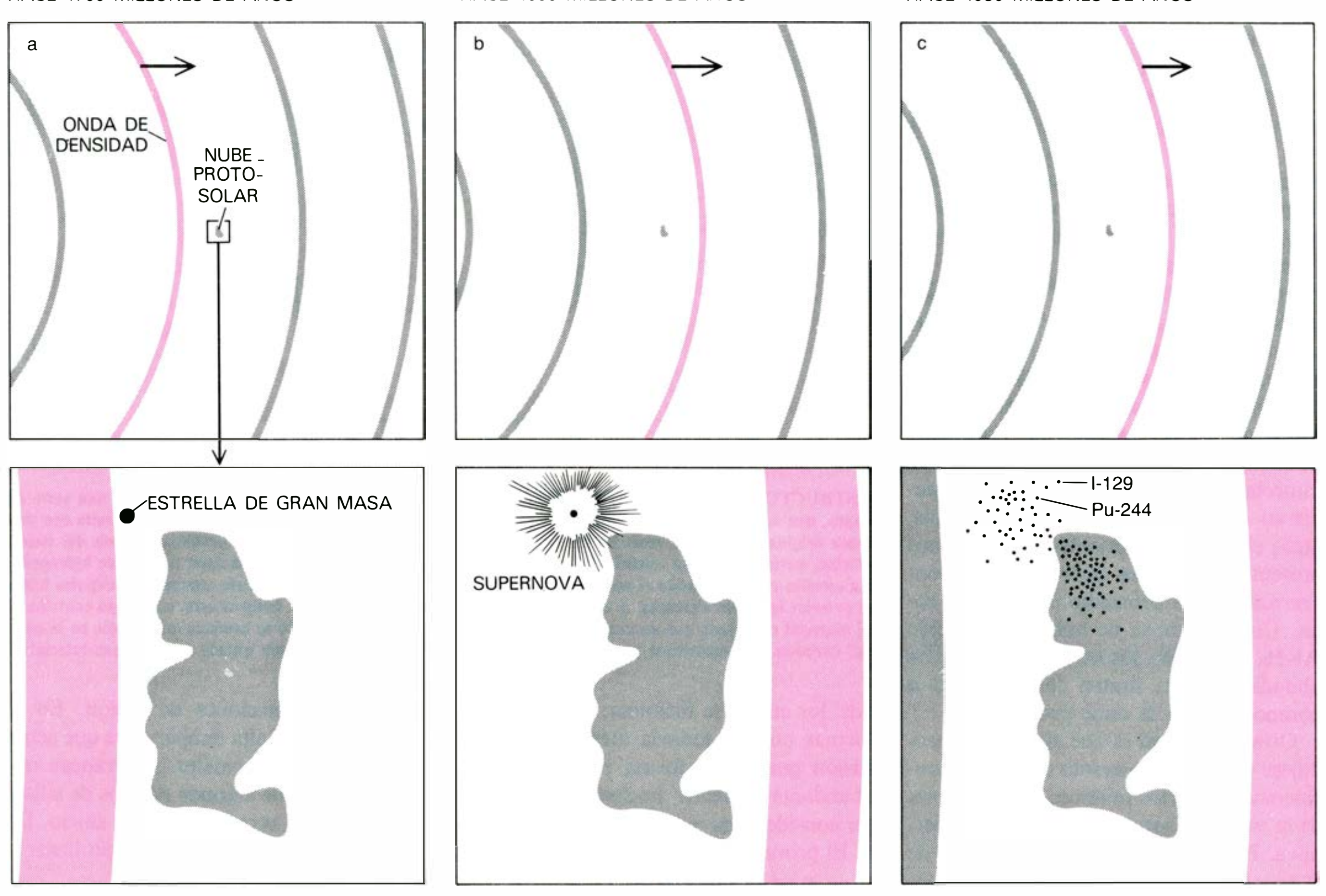
Una característica común a todas esas reacciones de fusión nuclear es que liberan energía. En todos los casos, el núcleo más grande es más estable que el menor a partir del cual se edificó, de manera que en la



ESTRUCTURA DE UNA SUPERNOVA inmediatamente antes de que explote; consta de una serie de capas, que se miden aquí en unidades de masa solar. La estructura es la prevista para una estrella con una masa original igual a 22 veces la del Sol; en esta fase de su evolución hubiera perdido, a través del viento estelar, parte de la masa inicial, pero se desconoce la verdadera extensión de la capa externa de hidrógeno. La estrella explota cuando el núcleo de hierro y níquel se colapsa, desprendiendo energía que expulsa hacia el exterior las capas externas. A causa de las diferencias en composición y temperatura, cada capa contribuye al material expulsado con un conjunto característico de isótopos. El Al-26 se produce tal vez sólo en la capa del carbono. Los elementos pesados, desde el silicio al níquel, se producen quizás en las capas interiores.



SE PUEDE OBSERVAR LA FORMACION DE ESTRELLAS en la parte anterior de los restos en expansión de una supernova. La nebulosa brillante, que aparece obscura en esta imagen negativa, es la constelación Perro Mayor. En las regiones marcadas con flechas, donde evidentemente los productos expulsados de la supernova chocan con la materia del medio interestelar, se han descubierto concentraciones de estrellas jóvenes y brillantes. La condensación de las estrellas puede inducirse por ondas de choque procedentes de la supernova, y el Sol pudo haberse formado por medio de una interacción similar con la supernova causante de las anomalías de magnesio y oxígeno. Esta prueba de la formación de estrellas en la proximidad de los restos de una supernova fue descubierta por William Herbst y George E. Assousa, de la Carnegie.



ONDAS DE DENSIDAD asociadas a la estructura espiral de la galaxia. Estas ondas provocan episodios de formación de estrellas al atravesar una región dada de la galaxia una vez cada cien millones de años. Las anomalías isotópicas en los meteoritos parecen incluir indicios del paso de dos ondas de densidad sucesivas. Hace unos 4700 millones de años (*a*), una onda produjo estrellas, incluyendo al menos una de enorme masa, en la vecindad de la nube de gas y polvo que había de dar origen al sistema solar. Dicha estrella se convirtió en una supernova (*b*) e inyectó cierta cantidad de nuevo material en la nube (*c*). Entre los isótopos inyectados se encontraban el iodo 129 (con una semivida de 17 millones de años) y el plutonio 244 (con una semivida de 82 millones de años),

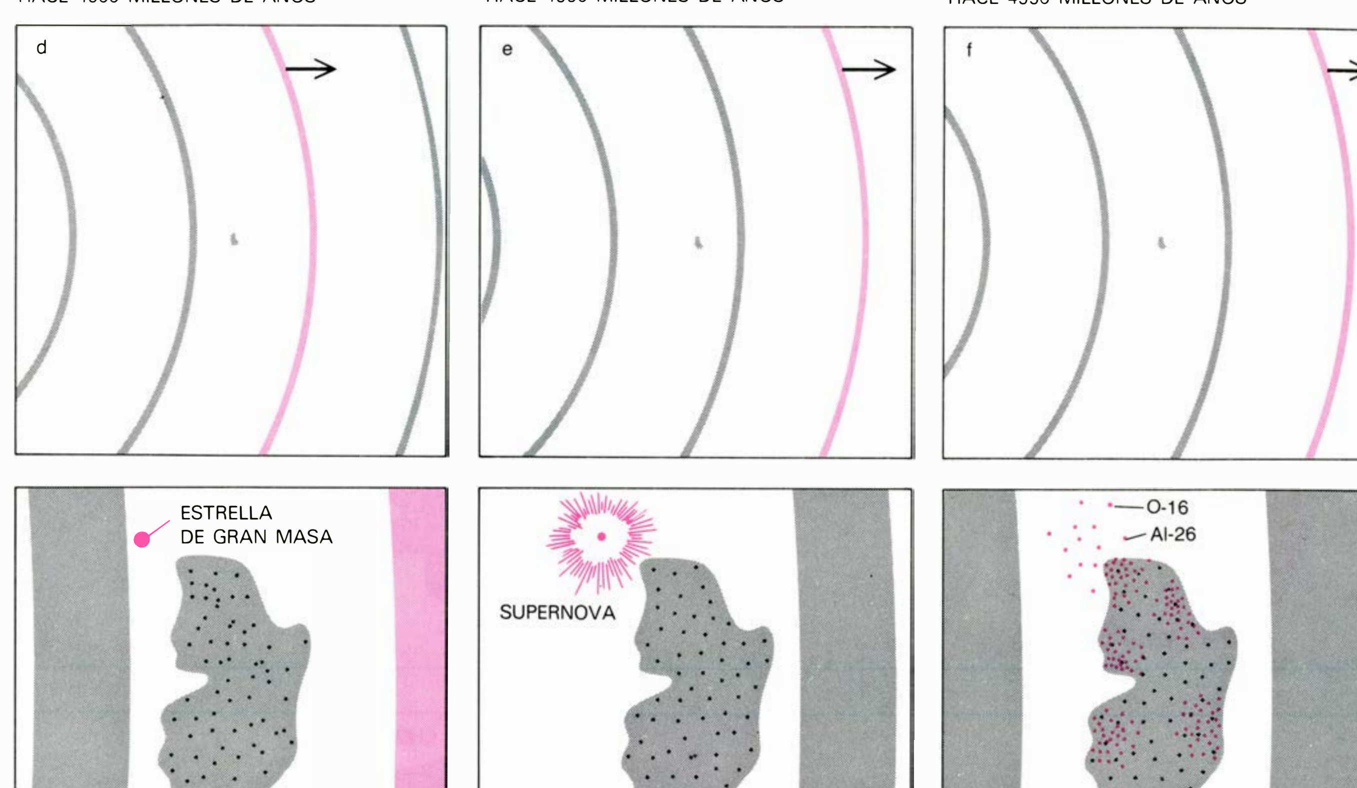
fusión se libera calor. La secuencia debe detenerse, sin embargo, en el hierro, porque éste es el más estable de todos los núcleos. La expulsión de este material recién creado desde el lugar donde se produce, en el centro de estrellas de gran masa, hacia el espacio interestelar, requiere la demolición de las estrellas en explosiones de la clase supernova. Además, muchos otros elementos entre el carbono y el hierro se elaboran solamente bajo las condiciones extremas que rigen en las propias explosiones de las supernovas.

Solamente las estrellas cuya masa sea mayor que unas seis veces la del Sol pueden transformarse en supernovas. Cuando una tal estrella ha agotado su reserva de combustible por fusiones nucleares, se convierte en estratificada. En el núcleo hay hierro y níquel, después una capa de silicio que también incluye indicios de magnesio y oxígeno, y a continuación una capa de oxígeno, otra de neón y otra de carbono, la cual también tiene algo de oxígeno, neón y magnesio. En la periferia están los elementos más ligeros: helio (con indicios de carbono 12) y una envoltura externa de hidrógeno.

La clase exacta de inestabilidad que produce que una tal estrella explote no se conoce bien. De alguna manera, la energía que mantiene la presión interna del núcleo de hierro y níquel se extrae de la región central y se transfiere a capas más periféricas. Ello deja en libertad al núcleo para colapsar bajo su propia atracción gravitatoria (para formar una estrella de neutrones o quizás un agujero negro), mientras que todas las capas exteriores son expelidas a gran velocidad. La explosión es tan enérgica que algunas supernovas lucen con más brillo que galaxias enteras durante un corto período de tiempo.

Durante la explosión, las capas expelidas se calientan a temperaturas de miles de millones de grados Kelvin, posibilitando así reacciones nucleares que quizá no tienen lugar en ninguna otra parte del universo (excepto en aceleradores de partículas especialmente proyectados). Existe un intenso flujo de partículas, incluyendo neutrones, protones y núcleos de helio. Los neutrones importan, en especial, en la síntesis de los elementos pesados. Un núcleo de hierro, por ejemplo, puede absorber muchos neutrones, creando un isótopo de hierro pesado y extremadamente inestable que se desintegra por emisión beta. Para los elementos más pesados que el plomo y el bismuto y para algunos elementos más ligeros, los neutrones deben absorberse en sucesión rápida.

El aluminio 26 se elabora en la capa de carbono de las supernovas. El magnesio 24 se acumula en esta región mediante la fusión de pares de átomos de carbono, y los isótopos más pesados de magnesio se forman entonces por absorción de neutrones. Puede producirse algo de Al-26 cuando un núcleo de Mg-26 absorbe un protón e inmediatamente emite un neutrón. Otro método de producción comienza con Mg-24, que puede absorber un neutrón (convirtiéndose en Mg-25) y luego un protón (produciendo Al-26). En la proximidad de una estrella inmediatamente antes de una explosión del tipo supernova, son posibles muchas reacciones similares con el mismo producto final. Ninguna de ellas constituye un importante camino nucleosintético, pero eso era de esperar ya que ni el Al-26 ni siquiera el Al-27 son núcleos que abunden de una manera destacada.



una parte de los cuales sobrevivieron hasta que se formó el sistema solar. Esa parte se puede detectar en los meteoritos por medio de los productos de desintegración del iodo 129 y el plutonio 244, que comprenden isótopos del xenón. Hace unos 4600 millones de años otra onda atravesó la misma región y dio nacimiento a otra generación de estrellas de gran masa (*d*), incluyendo una que estaba probablemente a menos de 60 años-luz de la nube protosolar. Cuando esta estrella explotó (*e*), añadió a la nube otro conjunto de isótopos (*f*). Por encontrarse entre éstos el aluminio 26, de vida corta, se sabe que la nube debió condensarse poco después, y la condensación pudo haber sido inducida por la explosión de la supernova. En la ilustración, O-16 es un isótopo del oxígeno.

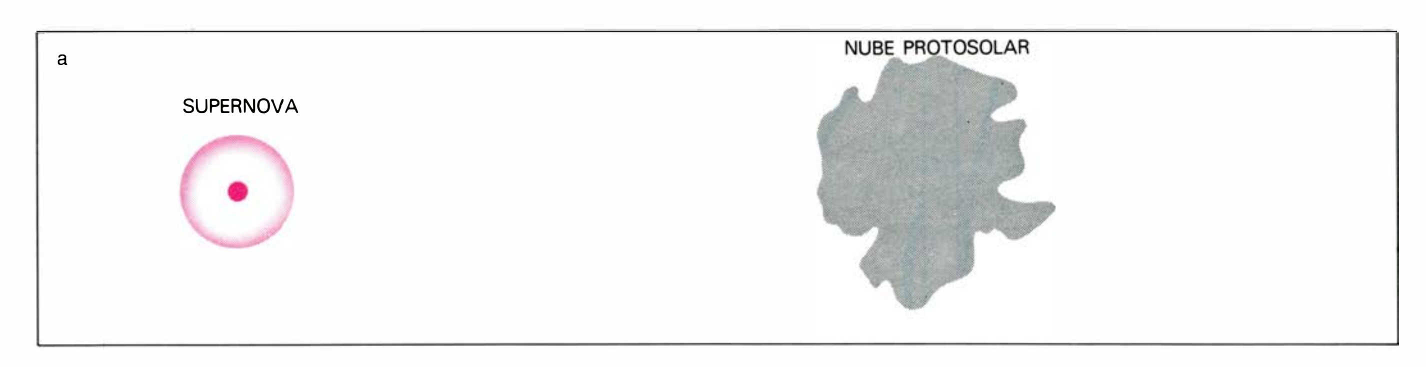
Esta clase de mecanismos, que actúan sólo en las supernovas, son los responsables de la producción de la mayorías de los elementos que no se encuentran en la secuencia directa de núcleos de helio. Hay unas pocas excepciones, sin embargo, entre las más importantes de las cuales se encuentran el oxígeno 17 y el oxígeno 18. Esos isótopos se pueden formar en supernovas (como resultado de captura de neutrones por el O-16, por ejemplo), pero no sobreviven el tiempo suficiente para penetrar en el medio interestelar. El oxígeno 17 es un núcleo frágil, y a las temperaturas de la supernova se desintegra a una rápidez mayor de lo que se forma. El oxígeno 18 es ligeramente más resistente, pero absorbe con facilidad un núcleo de helio y la mayor parte de él se convierte de esta manera en neón 22.

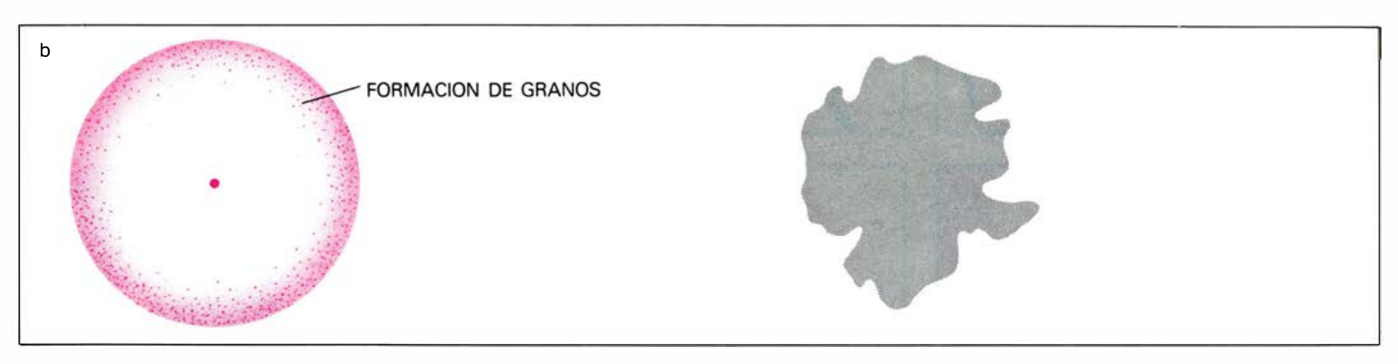
La más probable fuente de los isótopos pesados de oxígeno que se hallan en el sistema solar es la envoltura exterior de las estrellas en envejecimiento llamadas gigantes rojas. El O-17 y el O-18 se producen allí (y en otras estrellas) como productos intermedios de la fusión catalítica de hidrógeno en el ciclo del carbono-nitrógeno-oxígeno. En una estrella gigante roja, el O-17 y el O-18 podrían no ser destruidos por exposición a altas temperaturas, de modo que algunos de esos átomos podrían escapar en el viento estelar que está siendo continuamente emitido por la atmósfera de la estrella. Este mecanismo es incapaz de dar cuenta de la producción de grandes cantidades de O-17 u O-18, pero ambos isótopos se encuentran sólo en muy pequeñas cantidades. Probablemente todo el O-17 de la nube protosolar se produjo así, y lo mismo debió suceder con la mayor parte del O-18.

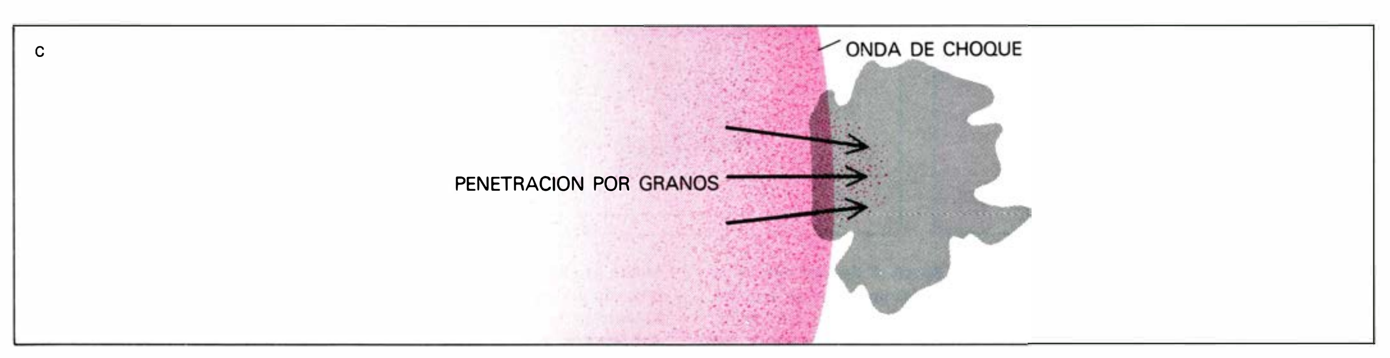
Es notable que tanto las anomalías de oxígeno como las de magnesio se puedan explicar por la introducción de materia procedente de una sola capa de las supernovas: la capa del carbono. Esta es la más externa de las zonas que contienen elementos más pesados que el hidrógeno y el helio. El Al-26 se produce solamente en la capa del carbono. También se encuentra el O-16, pero no el O-17, ni probablemente tampoco el O-18.

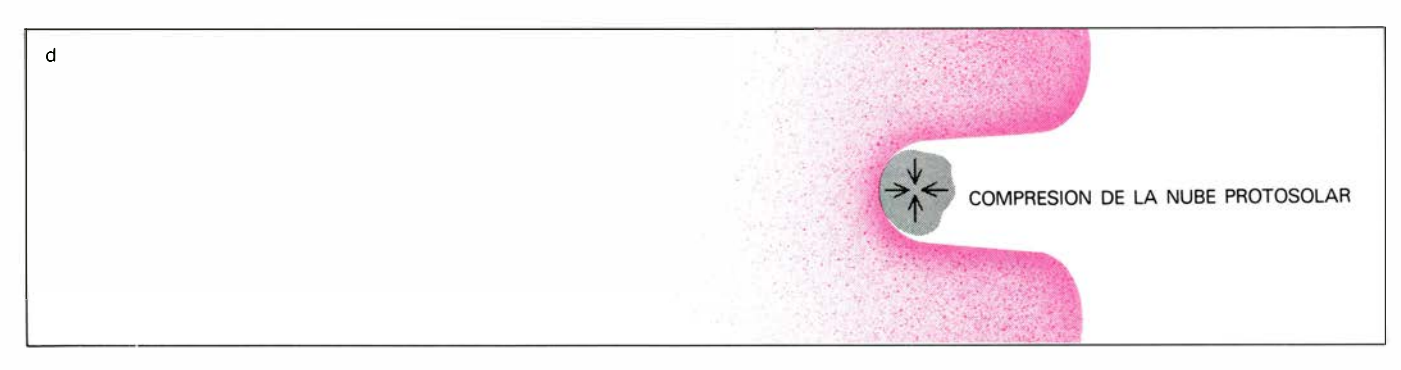
Antes del descubrimiento del exceso de Mg-26 en el material de Allende, se habían interpretado algunas otras anomalías isótopicas como productos de desintegración de especies radiactivas extintas. Hace más de diez años, John H. Reynolds y sus colegas de la Universidad de California en Berkeley hallaron que en algunos meteoritos había trazas del gas inerte xenón con una composición isotópica desusada. El exceso de Xe-129 en estos meteoritos se demostró ser un producto de desintegración del iodo 129. Se encontraban además cuatro isótopos más pesados del xenón (Xe-131, Xe-132, Xe-134 y Xe-136) en proporciones que indicaban que se habían formado por fisión espontánea del plutonio 244.

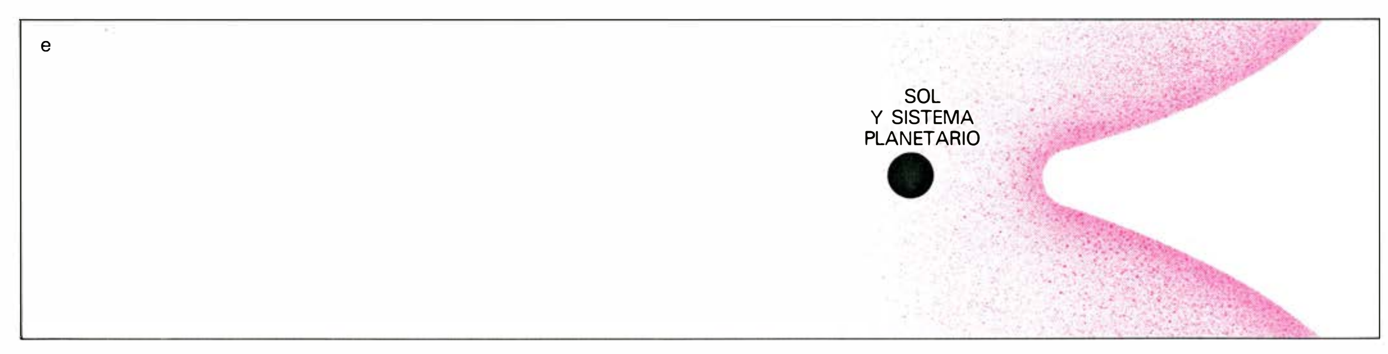
Tanto el iodo como el plutonio se producen en supernovas, pero en ninguna otra parte. De las semividas de los isótopos padres (diecisiete millones de años para el I-129 y 82 millones de años para el Pu-244), se pudo deducir el tiempo que había transcurrido entre la formación del iodo y el plutonio y el encierro del xenón en la materia sólida de los meteoritos. Los isótopos radiactivos se produjeron unos 100 millones de años antes de que los meteoritos se condensasen. Era obvio que no











LA ONDA DE CHOQUE de la explosión de una supernova próxima pudo haber proporcionado las fuerzas de compresión necesarias para convertir una nube difusa de gas y polvo en el Sol y su sistema planetario. Al principio, la nube de materia expulsada por la supernova debió tener la forma de gas (*a*), pero hay razones para creer que se condensaron relativamente pronto ciertos gránulos constituidos por minerales refractarios (*b*). Cuando los restos en expansión de la supernova encontraron a la nube protosolar (*c*), algunos de los gránulos pudieron penetrar en la nube como esquirlas, llevando con ellos las anomalías isotópicas que hoy se observan en los meteoritos. La mayor parte del gas, por otro lado, habría rodeado la parte exterior de la nube, comprimiéndola (*d*) hasta que excedió una densidad crítica y se colapsó bajo su propia gravitación para formar una estrella (*e*). (La ilustración es de George V. Kelvin.)

podían proceder del mismo suceso que dio lugar a las anomalías del magnesio.

El intervalo de 100 millones de años entre las dos supernovas tiene una explicación llamativa aunque hipotética. En el disco de nuestra galaxia, los episodios importantes de formación de estrellas están asociados con los brazos espirales, que aparecen brillantes precisamente porque las estrellas jóvenes están concentradas allí. Se cree que los brazos espirales son ondas de densidad que se propagan alrededor del centro de la galaxia de manera análoga a las compresiones y enrarecimientos de una onda sonora. Las compresiones inducen el que las nubes de polvo y gas se concentren en estrellas, y las de mayor masa entre ellas viven sólo algunos millones de años antes de explotar como supernovas. Lo que es más significativo: una onda de densidad pasa a través de una región dada en la galaxia aproximadamente una vez cada 100 millones de años, y, por tanto, deben esperarse contribuciones importantes de elementos pesados al gas interestelar el mismo intervalo de tiempo, también aproximadamente.

Se puede hacer una objección obvia a esta hipótesis. No es sorprendente que el O-16 y el Al-26 producidos por la primitiva supernova no dejaran trazas en los meteoritos: el O-16 se habría mezclado por completo con todo el resto de oxígeno presente, y el Al-26 se habría desintegrado antes de que el sistema solar se condensase. Es preocupante, sin embargo, que no se haya encontrado indicación del I-129 o Pu-244 procedentes de la supernova más tardía. Puede que tal ausencia de indicios no se explique nunca de un modo satisfactorio, pero hay dos posibilidades que merecen considerarse. En primer lugar, debe indicarse que el I-129 y el Pu-244 se producen sólo por la captura rápida de muchos neutrones. Este proceso no se conoce bien y es muy verosímil que no se produzca en todas las supernovas. Quizá ni el I-129 ni el Pu-244 entraron a formar parte de la nube protosolar después de la supernova más reciente, por la sencilla razón de que no se produjo ninguno de ellos en tal explosión. Otra posibilidad es que sólo las zonas más externas de la supernova tardía pudieron penetrar en la nube protosolar. Los elementos más pesados pueden proceder de capas más profundas.

Además de las especies radiactivas extintas, se han encontrado otras anomalías entre los isótopos del xenón y se conocen varios más en otros dos gases nobles: el kriptón y el neón. Estos descubrimientos no se han explicado completamente, aunque la anomalía del neón (que fue descubierta en 1972 por David Black, del Centro de Investigación Ames de la Administración Nacional de Aeronaútica y del Espacio, NASA), así como las otras, puede requerir una transformación nuclear fuera del sistema solar.

Se han detectado en el material de Allende otras relaciones isotópicas que no son corrientes. En los minerales de Allende, los excesos de Mg-26 y de O-16 no están estrechamente correlacionados: un grano mineral que tiene un exceso grande de O-16 no tendrá necesariamente otro grande de Mg-26, ya que los excesos de Mg-26 se encuentran sólo en minerales ricos en aluminio. En dos muestras de Allende, sin embargo, ambas anomalías son grandes y tienen un carácter algo distinto de la perturbación isotópica en el resto del mineral. Parte del efecto es una consecuencia de procesos nucleares, pero ha habido también fraccionamiento químico en gran escala: unas 10 veces más que en otras muestras. En una de ellas se han medido relaciones isotópicas desusadas, además de en el oxígeno y magnesio, en el silicio, calcio, estroncio, bario, neodimio y samario. De hecho, todos los elementos examinados en esa muestra han presentado relaciones isotópicas anómalas. La explicación más razonable es que la muestra recibió una proporción excepcionalmente grande de materia procedente de la supernova reciente, que se condensó antes de que pudiera mezclarse con otros componentes de la nube. En particular, la muestra puede haber sido contaminada con una proporción abundante de materia de la capa del carbono de la supernova.

Dada la escasez de las supernovas, habría sido una coincidencia altamente improbable que una de ellas explotase en el momento y lugar precisos en que el sistema solar estaba a punto de formarse. La estrecha asociación entre los dos sucesos quizá no fuera una pura coincidencia: por el contrario, tanto argumentos teóricos como la observación de restos de supernovas distantes sugieren que la explosión pudo haber sido parte del mecanismo de la creación del sistema solar.

Los cálculos teóricos muestran que una nube aislada de polvo y gas, suficientemente difusa y con una masa comparable a la del Sol no puede experimentar un colapso bajo la influencia de su propia gravitación a menos que exista algún estímulo exterior. Las fuerzas gravitatorias decrecen con el cuadrado de la distancia, de modo que, en una nube de gas y polvo fino, la fuerza de cohesión está determinada en gran parte por la densidad. A menos que la densidad exceda cierto valor crítico, basta la presión interna del gas para impedir la condensación de la nube. Pero una vez se alcanza la densidad crítica, el colapso es espontáneo. La necesidad de un suceso iniciador es un problema antiguo en la teoría de formación de las estrellas. La posibilidad de que la explosión de una estrella próxima suministrase la compresión necesaria se propuso hace más de 30 años; en particular, fue expuesta por Fred Hoyle y E. J. Öpik. Más recientemente, Paul Woodwards, del Laboratorio Lawrence Livermore de la Universidad de California, ha demostrado que las ondas de choque en expansión pueden aplicar una fuerza de compresión suficiente para crear una estrella. A medida que el frente de choque pasa a través de una nube difusa de gas, la nube queda envuelta y comprimida en un volumen menor. Estos cálculos generalizados de ondas de choque se pueden aplicar al caso específico de la onda de choque de una supernova que incide sobre una nube de gas interestelar.

William Herbst y George E. Assousa, de la Institución Carnegie de Washington, han tomado recientemente fotografías que parecen mostrar ese proceso. Tales fotografías son de nebulosas en expansión remanentes de supernovas en otros lugares de la galaxia. En el borde anterior de tales restos de supernovas, Herbst y Assousa han encontrado concentraciones de estrellas jóvenes y brillantes.

La distancia entre la nube protosolar y la supernova que provocó su colapso se puede estimar a partir de la energía necesaria para la compresión, de la dilución de las anomalías isotópicas observadas y del tamaño de las regiones donde se forman estrellas, incluyendo las observadas por Herbst y Assousa. La mejor estima que cumple con esas limitaciones establece que la supernova no estaba más allá de 60 años-luz de la nube, lo que la convertiría en una de las estrellas más próximas. No hay hoy esperanza de identificar el núcleo residual de la supernova. La propia explosión dio probablemente al núcleo una alta velocidad respecto al sistema solar, de modo que en la actualidad debe estar muy lejano. Además, desde entonces se han producido unas 40 a 50 ondas galácticas de densidad, cada una de las cuales añade una nueva generación de estrellas y supernovas.

La supervivencia de minerales que transportan isótopos que pueden adscribirse a una supernova particular es más plausible si tales isótopos entraron a formar parte de la nube protosolar en forma de granos sólidos que si fueron inyectados en ella en forma gaseosa. Los gases se mezclan con tal rapidez que las distintas relaciones isotópicas se igualarían quizás en menos de un millón de años. Steven Margolis, de la Universidad de Chicago, acaba de mostrar que los granos tienen también mayor probabilidad de penetrar la nube protosolar; los granos actúan como esquirlas, mientras que los

gases procedentes de la supernova tienden a rodear la nube. Una vez dentro de la nube, los granos tienden a reunirse, aumentando con ello la probabilidad de que pueda observarse una anomalía isotópica, aun cuando la masa total aportada por la supernova más reciente fuese pequeña en comparación con la masa del sistema solar.

Estos descubrimientos tienen sentido sólo en el caso de que en las explosiones de supernovas se formen granos reales. Margolis y Sydney Falk, también de la Universidad de Chicago, junto con James Lattimer, de la Universidad de Illinois, han hallado que la condensación de granos sólidos es previsible a partir de un modelo corriente de explosiones de supernovas. Los granos no se forman en la explosión misma, sino cuando la nebulosa se expande y enfría. Proceden de capas interiores, donde se encuentran los elementos más pesados, y algunos de los granos que se solidifican a temperatura más alta son precisamente los que tendrían mayor exceso de oxígeno 16. Tales minerales refractarios estarían entre los primeros en formarse y entre los más resistentes a una ulterior vaporización. No es razonable esperar que ninguno de los granos minerales hallados hoy en los meteoritos se produjeran en la supernova y se hayan conservado inalterados; las propias condritas carbonáceas han sufrido alguna vaporización, fusión o procesos químicos. En los granos modernos, sin embargo, el material de la supernova aporta una contribución importante.

Las relaciones isotópicas medidas se pueden combinar con otras indicaciones para establecer una explicación provisional de los sucesos que dieron lugar a la formación del sistema solar. Algunos aspectos de esta explicación son necesariamente especulativos, en particular la conclusión de que una supernova desencadenó el colapso de la nube protosolar. Las anomalías isotópicas no son concluyentes en este contexto y sólo demuestran que los dos acontecimientos se produjeron aproximadamente en el mismo tiempo y lugar. La hipótesis de la causalidad parece, sin embargo, más plausible que la de una mera asociación accidental.

En el tiempo en que se formó el sistema solar, la galaxia había estado desarrollándose durante al menos siete mil millones de años y quizás hasta quince mil millones de años, de modo que ya se habían producido muchas generaciones de estrellas. El polvo y gas que se encontraba en abundancia en el plano de la galaxia se había enriquecido en elementos más pesados por nucleosíntesis estelar y supernovas en esas generaciones primitivas. Puede suponerse que, como consecuencia de la mezcla de las contribuciones de muchas supernovas y de otras fuentes, la composición del polvo y gas era uniforme en toda la región en que se formó el sistema solar.

Hace unos 4700 millones de años pasó a través de esa región una onda de entre la serie de ondas de densidad asociadas a la estructura espiral de la galaxia. En la estela de la onda de densidad se formaron algunas estrellas, incluyendo unas pocas de masa muy grande que consumieron en seguida su combustible nuclear y explotaron como supernovas. Al menos una supernova de esta generación creó iodo 129 y plutonio 244, parte de los cuales penetraron en la nube protosolar. Sin duda, se produjo también aluminio 26, pero todo él se desintegró durante los varios millones de años siguientes. Los productos de desintegración de tales especies radiactivas de vida corta, junto con los isótopos estables procedentes de la supernova, se mezclaron con la nube y no dejaron traza.

Unos 100 millones de años más tarde, es decir, hace 4600 millones de años, pasó otra onda de densidad y se produjo otra generación de estrellas de gran masa. Esta vez, explotó una supernova a una distancia menor que unos 60 años-luz de la nube protosolar; de hecho, la supernova pudo haberse producido a expensas de otra parte de la misma nube. No se incorporó al sistema solar iodo ni plutonio procedente de esa supernova ni de ninguna otra de la misma generación, de modo que parece que la supernova próxima haya podido ser una de las de la primera generación. Sí alcanzaron la nube aluminio 26, oxígeno 16 y otros núcleos recién formados, junto con una onda de choque que los indujo a concentrarse en una estrella y un sistema planetario. En la mayor parte de los cuerpos del nuevo sistema solar, los procesos químicos borraron todo indicio de los productos de la supernova. Sólo en los cuerpos más pequeños y fríos, los meteoritos, se conservaron las relaciones isotópicas peculiares. Algunos de esos cuerpos se condensaron lo suficientemente pronto para que la radiactividad del aluminio 26 dejará trazas fósiles.

Las palabras nova y supernova se derivan del latín *stella nova*, "nueva estrella". Este era un término descriptivo en la época en que el fenómeno se percibía como la súbita aparición de una estrella donde antes no había ninguna visible, pero el nombre adquirió cierto grado de ironía cuando aprendimos que una supernova representa no la creación de una estrella, sino la muerte cataclísmica de ella. Ahora, sin embargo, parece que una supernova presidió al menos el nacimiento de una estrella: el Sol.

Los dibujos prehistóricos en el suelo del Perú

Tienen tal aspecto de destinados a ser vistos desde el aire, que han dado pie a muchas especulaciones fantasiosas. La investigación arqueológica indica quién los realizó y por qué fueron hechos

William H. Isbell

En los desiertos costeros del Perú están los restos de antiguos sistemas de irrigación que encerraban una fascinación particular para Paul Kosok, de la Universidad de Long Island, estudioso de la América del Sur precolombina. Hace unos 40 años, mientras trazaba el mapa de lo que él consideraba como tenues zanjas de irrigación, hechas a partir de los afluentes del río Grande, en la costa sur, Kosok se topó con algo totalmente inesperado: en el valle de uno de estos afluentes, el río Nazca, se encontró con la imagen gigante de un pájaro de más de 30 metros de largo, dibujada en silueta como si su destino fuese el de ser vista desde lo alto.

Las zanjas de regadío rara vez forman figuras de animales, si es que en alguna ocasión lo hacen. Kosok hizo averiguaciones entre sus colegas peruanos y se enteró de que muchos "dibujos de tierra" incluso mayores, de tipo geométrico, habían sido señalados en la misma comarca. Estos incluían líneas largas (llamadas ordinariamente "pistas"), zigzags, trapezoides y espirales. El descubrimiento por parte de Kosok de un dibujo de tierra seminaturalista constituía, sin embargo, una sorpresa.

Entre quienes habían estudiado las figuras geométricas se hallaba Toribio Mejía Xesspe, de la Universidad Nacional de San Marcos. En un artículo publicado en 1938 había sostenido que los geoglifos de la zona de Nazca no tenían nada que ver con el regadío, y que debían haber desempeñado algún cometido religioso en tiempos precolombinos. Las exploraciones continuaron. Una araña enorme fue el siguiente geoglifo seminaturalista que vino a sumarse a la lista y pronto se le añadieron otros. En la actualidad el número de figuras de este género pasa de los 30.

Antes de que Kosok abandonase el Perú, en 1941, imaginaba que las "pistas" pudieran representar, por lo menos, líneas de mira astronómicas. Como posibilidad digna de investigación, así se lo sugirió a María Reiche, estudiante de matemáticas y astronomía educada en Alemania y residente en Lima. Esta mujer se ha ocupado desde entonces del estudio de estos dibujos.

¿En qué pueden la arqueología y otras disciplinas asociadas contribuir a explicar fenómeno tan raro como los geoglifos peruanos? Tres preguntas que parecen difíciles de contestar surgen inmediatamente: ¿Cuándo fueron hechos los dibujos? ¿Quién los hizo? ¿Por qué los hicieron? En los años transcurridos desde que Kosok dirigió la atención del público en general hacia estas curiosidades, las dos primeras preguntas han encontrado respuestas que las dejan fuera de toda duda. En cuanto a la tercera pregunta, se pueden ofrecer cierto número de conjeturas razonables. Aquí voy a examinar los pasos que han conducido a las respuestas tanto como a las conjeturas.

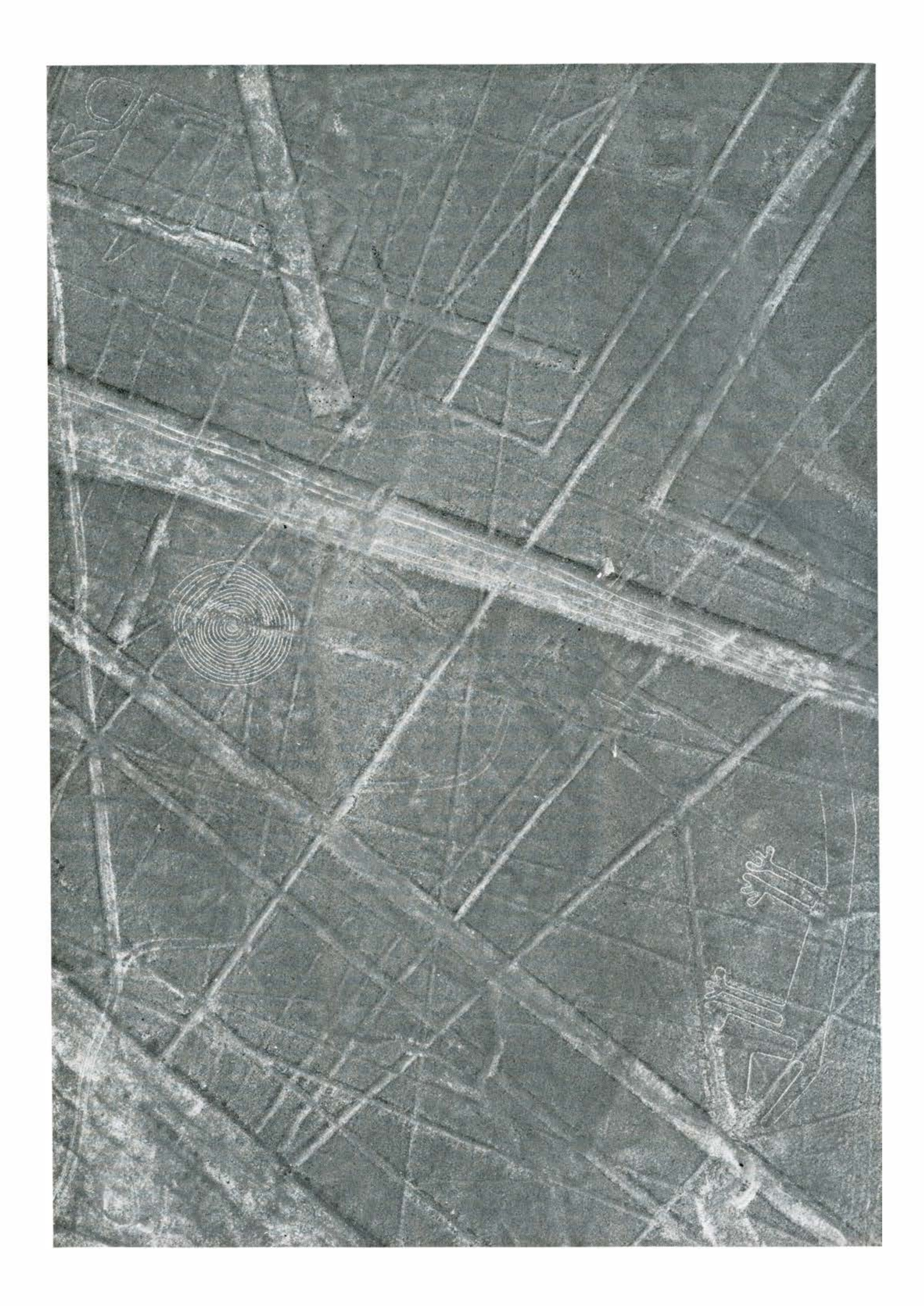
La costa meridional del Perú forma una hilada de colinas bajas que discurre en general de norte a sur. Entre esta elevación de la tierra ribereña y las estribaciones de los Andes, al este, se encuentra la depresión de una larga cuenca. Durante miles de años los arrastres de tierras más altas hacia el este y hacia el oeste han llevado los productos de la erosión hacia esa cuenca. La mayor parte del material erosionado es tierra fina de color claro. Ocasionales aluviones violentos aportaron también productos mayores de erosión: piedras cuyo tamaño oscila entre guijarros pequeños y grandes cantos rodados.

Donde el río Nazca baja de las estribaciones de la montaña para unirse al río Grande, estos milenios de relleno de erosión han formado una llanura ancha y horizontal. Los fuertes vientos del sur que soplaban a través del llano se llevaron una gran parte de la polvorienta superficie del suelo, dejando en pos de sí un "pavimento de desierto" compuesto de guijarros y cantos. Al comenzar la mañana, las piedras estaban mojadas de rocío, pero el resto del día se hallaban expuestas al sol abrasador del desierto. Como consecuencia, las piedras se oxidaron hasta adquirir su obscuro color pardo rojizo.

Este cambio de color redujo en la práctica la cantidad de erosión del viento en el desierto. El viento sur todavía sopla hoy en día, pero mientras las piedras se secan y se calientan al sol, su irradiación ayuda a mantener una capa superficial de aire caliente que actúa como amortiguador contra el viento. Por lo que se refiere a la erosión de la lluvia, la precipitación es tan escasa a lo largo de la costa sur del Perú, que sus efectos resultaron inapreciables durante 3000 años por lo menos.

Las circunstancias geológicas que han producido cientos de miles de kilómetros cuadrados de tablero natural en el sur del Perú en ninguna parte se ven mejor que en la planicie que domina el valle del encajonado río Nazca, una zona de unos 20 kilómetros de largo y 10 de ancho, conocida como Pampa Colorada (la Llanura Roja). Aquí si uno recoge del suelo una de las piedras pardas rojizas, queda expuesto a la vista el terreno claro que se halla debajo. Recogiendo una fila de pie-

TABLERO DEL DESIERTO de lo geoglifos, la Pampa Colorada, cerca de Nazca, tal como aparece en la foto aérea de la página opuesta. Se pueden ver varias clases de dibujos: líneas semejantes a pistas, trazadas en diferentes direcciones; zonas más anchas, limpiadas, en forma trapezoidal. Una espiral típica aparece en el centro, a la izquierda. A la derecha, abajo, se aprecia una figura de animal, quizás un lagarto.



dras aparece una línea de color claro. Así es como fueron hechos los geoglifos, a saber: desplazando con criterio selectivo el pavimento del desierto.

Los lienzos o tableros naturales son familiares a los arqueólogos en países muy distintos y alejados del Perú. En el sur de Inglaterra, donde una delgada capa de tierra vegetal cubre formaciones de yeso, los antiguos pobladores del país pusieron el yeso al desnudo para diseñar, entre otras figuras, las grandes imágenes de "caballos blancos" de los condados de Wiltshire y de Kent. El pavimento natural del desierto ha sido levantado también para formar figuras en los desiertos del sur de California, como han demostrado las exploraciones de Dean R. Snow, de la Universidad estatal de Nueva York en Albany. A decir verdad, si la erosión del viento y del agua fuesen tan someras en otras zonas desérticas del mundo como lo son en la costa meridional del Perú, el número total, de geoglifos conocidos hoy podría ser muchísimo mayor. Comoquiera que sea, todas las figuras de este género plantean un serio problema, a saber: ¿Cómo podemos determinar cuándo se realizó la obra?



COSTA OCCIDENTAL de América del Sur; está dominada por la abrupta elevación de los Andes. En Perú las estribaciones de la gran cordillera reducen la llanura costera a una estrecha franja de desierto que sólo es fértil allí donde los ríos hacen bajar al Pacífico las lluvias del altiplano. A lo largo de la costa meridional del Perú la lluvia casi no se conoce.

La cerámica prehistórica hallada en zonas de necrópolis del sur del Perú, por ejemplo el gran centro funerario de Cahuachi, tiene una decoración tan típica, que se le ha dado el nombre de Nazca, tomado de la ciudad moderna y del valle fluvial donde se han excavado muchos ejemplos de estas cerámicas. Las cerámicas de Nazca fueron hechas aproximadamente entre 200 a.C. y 600 d.C. Los estudiosos de la arqueología andina llaman a esta fase de la prehistoria del Perú el Período Intermedio Temprano.

Algunas de las vasijas de Nazca son figuras; asumen la forma de animales y de seres humanos. Otras llevan animales pintados: pescados, una ballena agresiva portadora de un trofeo, aves acuáticas, colibríes, reptiles, monos y llamas. Cuando se puso de manifiesto que un repertorio similar de figuras animales naturalistas se podía hallar entre los geoglifos de la Pampa Colorada, no se pudo eludir la conclusión de que una misma población prehistórica era responsable de ambas creaciones.

Las pruebas, sin embargo, eran meramente iconográficas; pero un argumento más directo estaba a punto de llegar. Cierta cantidad de cerámica rota se halla en la Pampa Colorada. Como la zona es demasiado árida para la colonización, parece razonable suponer que esos tiestos debieron haberse dejado por visitantes temporales del desierto: viajeros casuales, comerciantes, peregrinos (si es que los geoglifos tienen un carácter ritual) o cuadrillas de trabajadores (si la factura de los geoglifos se debió a un esfuerzo organizado). A comienzos de la década de 1970, Gerald S. Hawkins, del Smithsonian Astrophysical Observatory, reunió una cantidad de fragmentos de la Pampa Colorada y les pidió a Gordon R. Willey, de la Universidad de Harvard, y a John H. Rowe, de la Universidad de California en Berkeley, que se los identificasen. Willey y Rowe descubrieron que el 85 por ciento de los fragmentos correspondían a cerámica de Nazca. El 15 por ciento restante eran cerámicas de un período posterior: el comprendido entre 900 d.C. y 1400. Casi al mismo tiempo Rodger Ravines, del Instituto Nacional de Cultura del Perú, formaba una colección de fragmentos de yacimientos situados en la periferia de la Pampa Colorada: éstos resultaron ser exclusivamente tiestos de cerámica de Nazca. Los yacimientos periféricos dan la impresión de ser refugios temporales, quizá campamentos de cuadrillas de trabajo.

Estas pruebas de una identificación positiva de los usuarios de la cerámica de Nazca como visitantes de los geoglifos, unida a la presencia de unas mismas figuras de animales en éstos y en aquélla, resolvían de una vez para siempre el problema de quién había alterado con tanta habilidad el piso del desierto en la Pampa Colorada. Queda el hecho de que el intervalo entre 200 a.C. y 600 d.C. es de casi un milenio. ¿Sería posible señalar con mayor precisión cuándo tuvieron lugar esas actividades dentro del Período Intermedio Temprano? Una datación precisa de este género llegó a ser posible después de la Segunda Guerra Mundial merced al desarrollo de los análisis del carbono-14. Algunas de las líneas largas trazadas en la Pampa Colorada acaban donde hubo un poste de madera hincado en el suelo. En 1953 W. Duncan Strong, de la Universidad de Columbia, recogió una muestra de uno de esos postes para someterlo al análisis del carbono-14. El resultado indicaba que la madera pertenecía a un árbol que había sido cortado en 525 (± 80) d.C., en las postrimerías del Período Intermedio Temprano. Claro que en el mejor de los casos el resultado del análisis indica cuándo la línea larga asociada con este poste en concreto fue trazada (y quizá ni siquiera eso, si se piensa que la estaca pudo ser la segunda, la tercera o la trigésima que se puso para reemplazar a la original). Pese a todo, la fecha del carbono-14 cae dentro de los límites del período Nazca y, por tanto, da mayor peso a la otra prueba.

Ahora que las preguntas sobre el "qué" y el "cuándo" han sido contestadas, es el momento de considerar la cuestión del "por qué". Una vez más, la arqueología y disciplinas afines, especialmente la etnología, nos ofrecen útiles indicaciones. Por ejemplo, puede parecer raro que una población cualquiera dedique una cantidad significativa de energía a construir manifestaciones que se ven mejor desde el aire que desde el suelo. A decir verdad, el pueblo que hizo las figuras de yeso en Inglaterra las situó en laderas donde son claramente visibles. Pero, por otra parte, las manifestaciones del Nazca cuentan con paralelos. Entre las ambiciosas obras de tierra realizadas por los fieles del culto de Hopewell, de la Norteamérica precolombina, hay cierta cantidad de figuras que serían más fáciles de detectar desde el aire que desde el suelo. El gran túmulo serpentiforme de Ohio constituye un ejemplo sobresaliente. Otro tanto puede decirse de las figuras del desierto de California documentadas por Snow. Dado, pues, que este

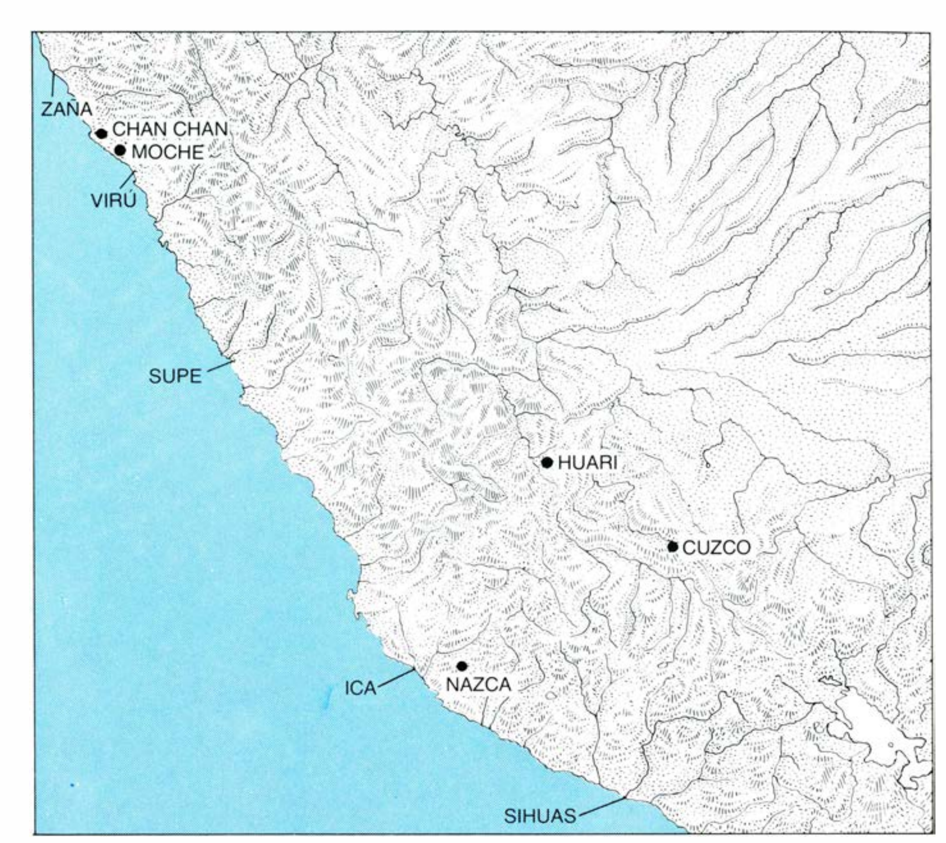
carácter de los geoglifos del Nazca no es único, parece innecesario suponer que los dibujos fueron hechos para ser vistos desde el aire, aun cuando pudiéramos imaginar cómo se hacían.

¿Qué decir del costo de la energía que había que emplear para mover las piedras y crear un geoglifo? Aquí vuelve la arqueología a darnos una pista. En otras zonas del Perú, durante el Período Intermedio Temprano, las poblaciones locales de granjeros emprendían trabajos colectivos de proporciones asombrosas. Por ejemplo, la mayor pirámide prehistórica del Perú está situada en el valle del río Moche. El enorme basamento del templo, el Huaca del Sol, requirió la manufactura de 140 millones de adobes. Los estudios de los métodos de construcción, y de los adobes mismos, indican que el trabajo fue llevado a cabo por obreros no especializados. Cada cuadrilla, probablemente compuesta de hombres reclutados en una sola región, manufacturaba sus propios adobes, los transportaba al emplazamiento del edificio y allí levantaba uno o más de los pilares de adobe que componían la pirámide. El trabajo de cada cuadrilla todavía se puede identificar por las marcas distintivas de las diferentes remesas de adobe.

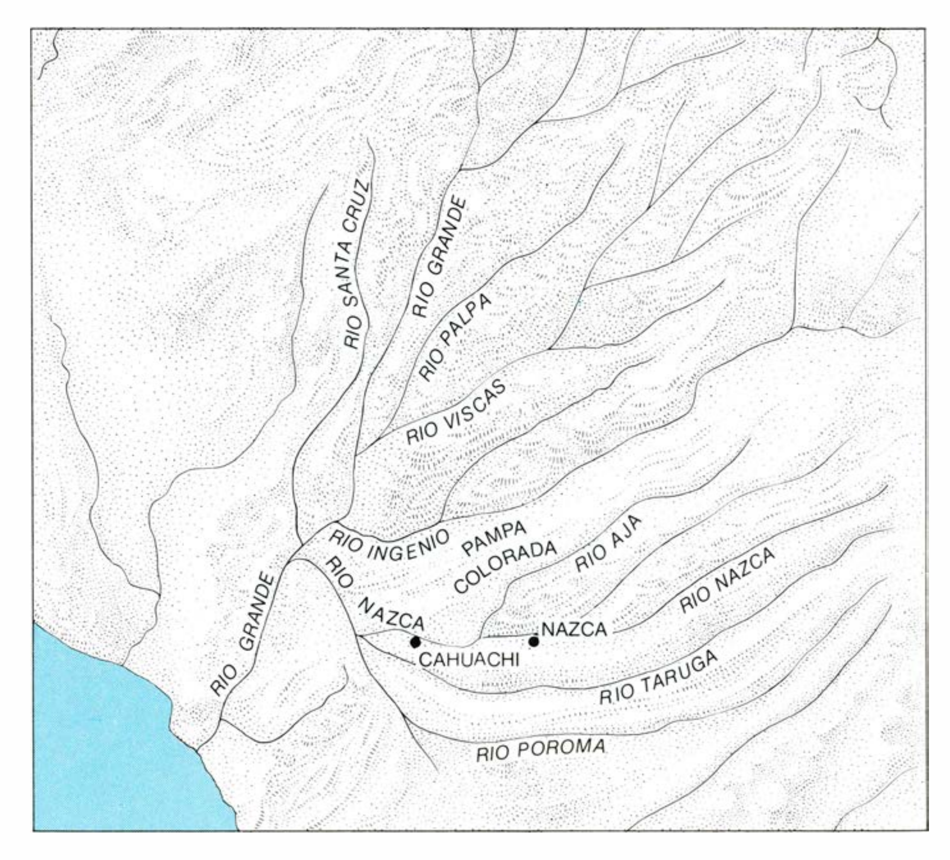
Las construcciones piramidales, y otros edificios monumentales del Período Intermedio Temprano en el norte y en el centro del Perú, representan evidentemente el apogeo de una época precolombina de la arquitectura ceremonial andina. Algunas plataformas de templos se encuentran también a lo largo de la costa meridional del Perú, pero son modestas en comparación con los enormes edificios de adobe que se levantaban en el norte. Y al revés, se hicieron geoglifos en sitios de los desiertos de las costas del norte y del centro del Perú, pero que son modestos en comparación con los módulos del Nazca.

Quienes no estén familiarizados con los geoglifos pueden encontrar dificultad para comparar el esfuerzo que su ejecución comporta con el esfuerzo necesario para hacer y colocar en su sitio los adobes. La razón es que las discusiones se han centrado hasta ahora, casi siempre, en las figuras de animales. En términos del tonelaje de piedra desplazada, las figuraciones de animales sólo representan una fracción del total de la energía invertida. Por ejemplo, un animal puede estar diseñado por medio de una sola línea continua; sin embargo, muchas de las líneas arrancan de un trapezoide o de un rectángulo situado a cierta distancia y al que luego vuelven.

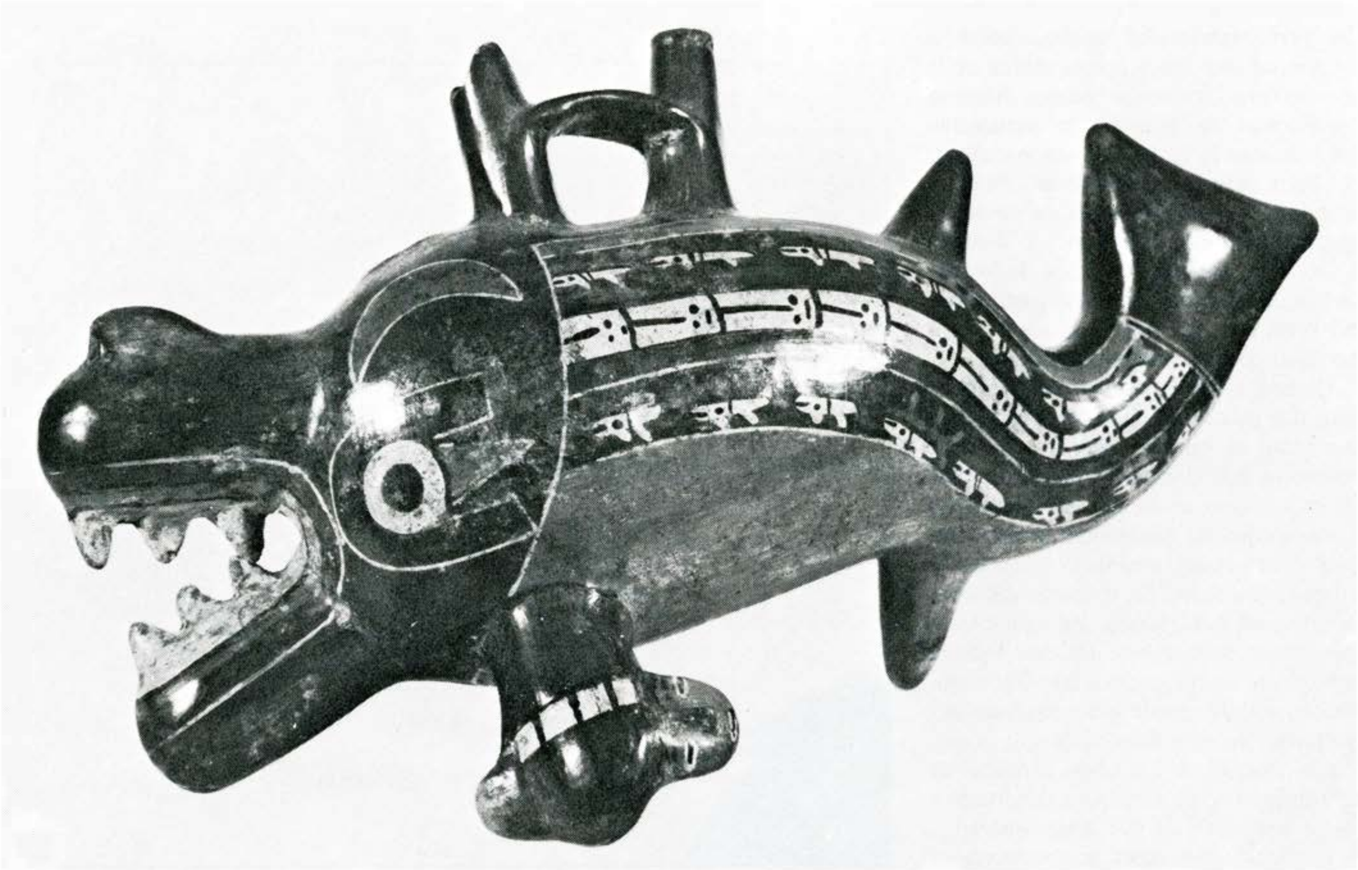
La figura geométrica que proporciona el punto de origen puede tener más de un kilómetro de longitud. La misma figura puede también constituir el punto de arranque de una serie de zigzags, o de una



LOS GEOGLIFOS han sido localizados en los desiertos costeros del Perú desde el valle del Virú, en el norte, hasta el valle de Sihuas en el sur. Se encuentran en las cercanías de Supe, Lima e Ica y se han señalado, pero no confirmado, en el valle de Zaña. Abundan mucho en las proximidades de Nazca, donde han aparecido tumbas antiguas que contienen una cerámica característica. Alan D. Iselin diseñó estas ilustraciones.



EL ALTIPLANO DEL DESIERTO al norte del río Nazca, la Pampa Colorada, mide unos 50 kilómetros de largo por 25 de ancho. Encierra la mayor concentración de geoglifos del Perú. Las cercanas ruinas de Cahuachi, una estación funeraria del Período Intermedio Temprano peruano (200 a.C. - 600 d.C.) constituyen uno de los numerosos viveros de la cerámica del estilo de Nazca, que es autóctona en esta parte del Perú.



UNA BALLENA AGRESIVA es el tema de este geoglifo de la Pampa Colorada (*arriba*). El objeto circular que se ve colgando bajo la mandíbula de la ballena es una cabeza humana, trofeo de guerra representado a menudo en los dibujos de la cerámica peruana y en otras decoraciones de la época precolombina. El vaso plástico (*abajo*) representa también una ballena homicida, acompañada de la cabeza - trofeo. La figurita, excavada cerca de la Pampa Colorada, pertenece al Período Intermedio Temprano de la prehistoria peruana. Lo reproducimos aquí por gentileza de nuestro colega peruano Wilfredo Loayza.

sola línea que recorra un trecho recto de desierto de ocho kilómetros o más, antes de terminar en una estaca o en un montón de piedras. Solamente si uno tuviese que medir el esfuerzo requerido para reproducir un geoglifo figurativo hoy, sería posible calcular el total de trabajo prehistórico invertido en la Pampa Colorada de una forma que no fuese la más general. No obstante, parece razonable suponer que el esfuerzo empleado en los geoglifos del Nazca equivale a una aplicación de energía comparable *grosso modo* con la que levantó los edificios monumentales de adobe en el norte.

Fuesen o no fuesen iguales los dos esfuerzos, ambos parecen haber cumplido funciones económicas similares. Estas funciones están relacionadas con el reclutamiento de trabajo comunal para obras públicas. Los controles llevados a cabo a escala nacional por los incas, en la época de la conquista del Perú por Pizarro, son bien conocidos. Menos divulgado está el hecho de que los más antiguos precursores del imperialismo inca tenían un sistema parecido de centralización. Mientras que la capital de los incas radicaba en Cuzco, la del imperio más antiguo se hallaba en Huari.

Centros administrativos provinciales, enlazados todos con Huari, reunían y almacenaban grandes cantidades de alimentos y bienes producidos en sus áreas rurales. Si ocurrían variaciones locales o regionales en la producción de la agricultura o de la industria de cabaña, la economía general se equilibraba mediante aportaciones a los almacenes centralizados, o por medio de intercambios entre los centros provinciales. Esta uniformidad económica permitió a la población del Perú prehistórico incrementarse y luego estabilizarse a un nivel más alto que el del período anterior de autonomía regional.

Durante el Período Intermedio Temprano la situación distaba mucho de la descrita. El trabajo comunal estaba sometido a algún tipo de control, como lo demuestra la construcción de los grandes basamentos de adobe de los templos. Sin embargo, no hay la menor señal de que la población de la época construyese edificios administrativos ni medios de almacenaje como los que fueron característicos de los regímenes imperiales de Huari y Cuzco. Esto sugiere que cada región estaba sometida a fluctuaciones económicas locales y a los cambios de población relacionados directamente con aquéllas.

Considérese el efecto de tales fluctuaciones. Una serie de años agrícolas buenos en una región hubiese dado lugar, en ausencia de algún mecanismo local de control, a un aumento de población. Seguidamente, una serie de años malos hubiera sido desastrosa para la población incrementada. La forma de evitar esta forma de reacción a las fluctuaciones económicas es impedir que el excedente de los años buenos dé pábulo a un aumento de población. Una nivelación artificial de la economía inhibirá el crecimiento de población, manteniendo el total por debajo de la capacidad máxima de sostén de la economía regional.

¿Cómo poner en práctica esta nivelación? Uno de los procedimientos consiste en fomentar la ocupación común en actividades ceremoniales que exijan una gran inversión de trabajo. En la época de Nazca

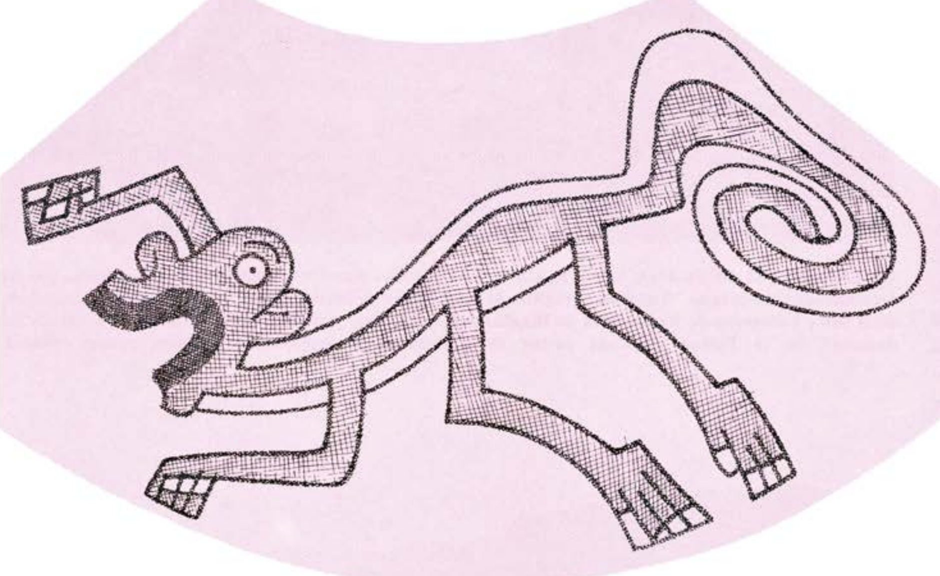


LA ESTUDIOSA PERUANA María Reiche (*a la izquierda*) **muestra los métodos de prospección que ha empleado durante más de 30 años en el registro de los geoglifos de Nazca. Su público es Patricia J. Knobloch, de la State University de Nueva York en Binghamton, arqueólogo visitante. Las piedras del "pavimento del desierto" en la Pampa Colorada varían de tamaño desde guijarros a grandes cantos rodados.**



ESTE DIBUJO EN ZIGZAG de la Pampa Colorada fue hecho, como lo fueron todos los geoglifos, retirando las piedras que cubren el suelo del desierto y poniendo al descubierto la tierra de color más claro que se encuentra debajo. Aquí algunas de las piedras desplazadas fueron apiladas en los pliegues de la figura.



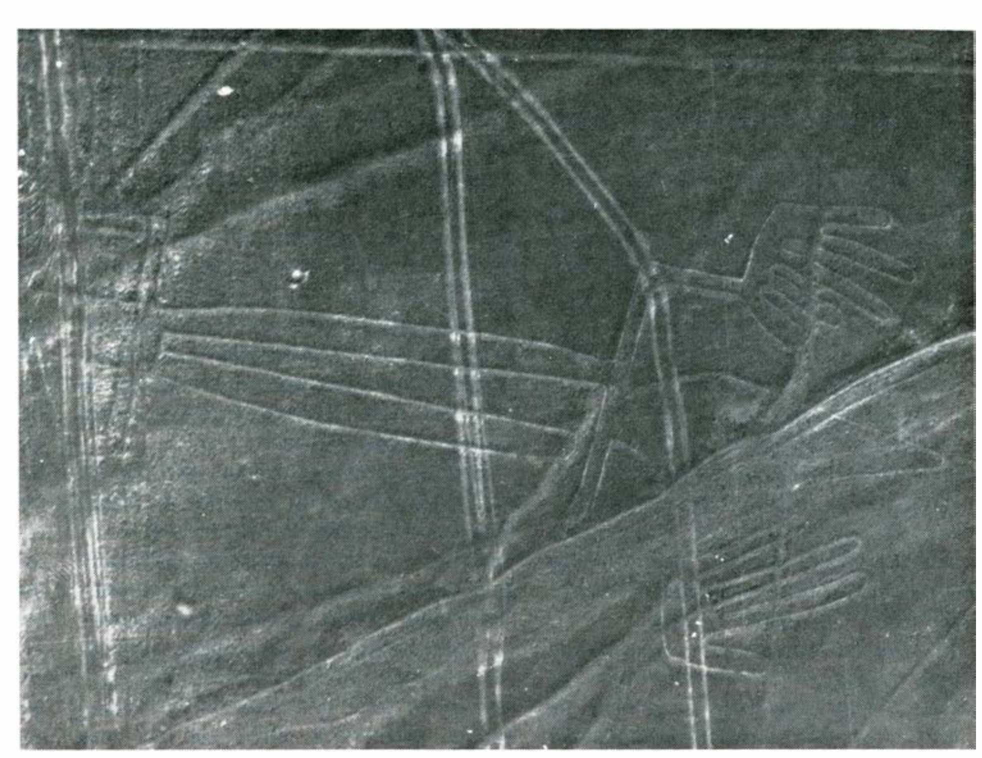
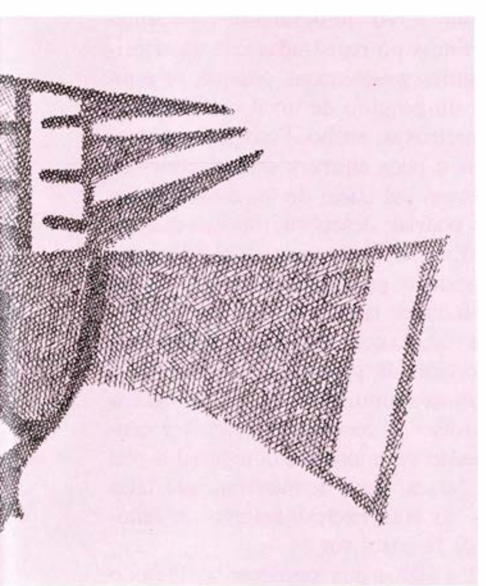
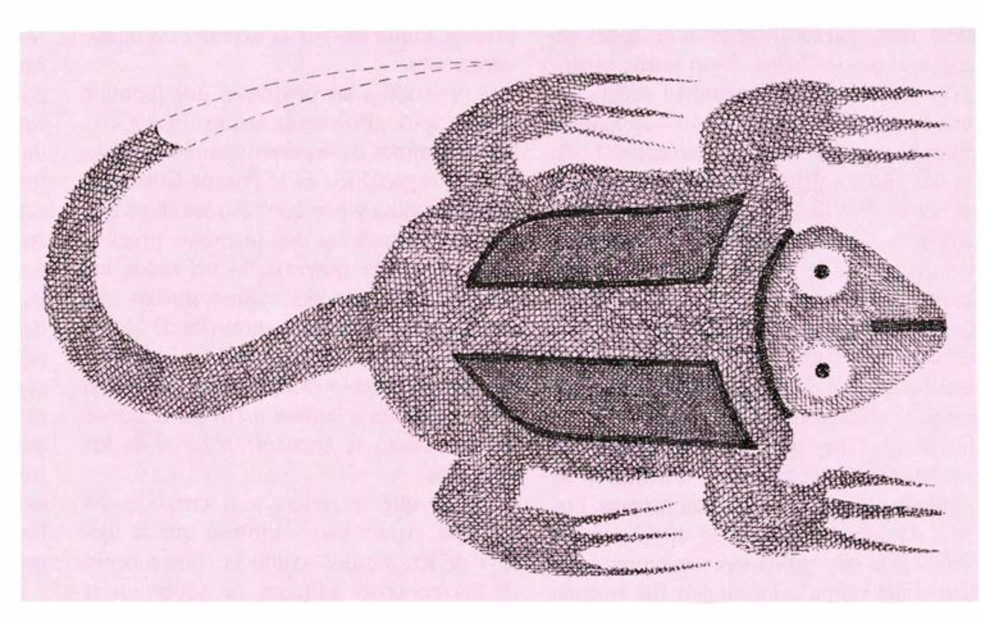



TRES FIGURAS DE GEOGLIFOS (*arriba*), comparadas con representaciones de los mismos animales que aparecen en la cerámica de Nazca (*abajo*). El de la izquierda es un mono estilizado con la cola exagerada. El dibujo de un mono se encuentra en un vaso de doble pitorro fotografiado por Loayza. La figura del pájaro del centro puede representar un rabihorcado; estos pájaros tienen un conspicuo buche en la garganta. El dibujo de un pájaro similar aparece

los tipos de almacenamiento centralizado de alimentos característicos de los regímenes de Huari y de Cuzco no existirían todavía, y los excedentes estarían diseminados no a nivel de provincia o de estado, sino local o incluso familiar. Si una región autónoma sufría unos años malos, el recurso a estos almacenes descentralizados ayudaría a remediar la situación. Si, por el contrario, la región disfrutaba de años buenos, los repletos almacenes particulares podían drenarse para sostener los equipos de trabajadores ocupados en actividades ceremoniales como la construcción de basamentos de templos o el diseño de geoglifos. En ambos casos la población permanecería relativamente estable.

Vista bajo esta luz, la función básica (o la ventaja selectiva, si se prefiere la terminología de la evolución darwinista) de llevar a cabo los geoglifos no tiene nada que ver con que si se veían desde el suelo o desde arriba, o incluso con que no se les viese desde ningún lado. Su función estriba en el hecho de que sociedades con un mecanismo cultural para invertir excedentes imprevistos en actividades ceremoniales gozan de una ventaja selectiva sobre sociedades carentes de tal mecanismo. Las primeras regulan su población, mientras que las sociedades que no lo hacen están condenadas a ciclos de "prospera" y "revienta".

¿Quizá los jefes de estas sociedades regionales del Período Intermedio Temprano de la prehistoria peruana tenían conciencia clara de conceptos tan complejos como el de la colocación de re-

en un vaso de Nazca en la colección del Putnam Museum de Davenport, Iowa. La figura como de lagarto, a la derecha, tiene una larga cola que no aparece en esta foto aérea, pero sus patas traseras no estaban diseñadas con el mismo detalle que las delanteras. El dibujo de un lagarto semejante aparece en un vaso de Nazca de la colección del Museo Robert H. Lowie de Antropología, que está adscrito a las dependencias de la Universidad de California en Berkeley.

cursos excedentes para impedir un exceso de población? Poco importa que fueran o no fueran conscientes de ello: el caso es que quienes poseían tal sistema acababan por reemplazar a sus vecinos y, por tanto, el proceder ceremonial coronado por el éxito se perpetuaba. A nivel de conciencia esta continuidad pudo haber estado fundada en un sistema de toma de decisiones no más profundo que la razón de que "así es como siempre lo hemos hecho".

Gran parte del comportamiento humano se basa en este tipo de "costumbre", simple continuidad de esquemas de comportamiento que tuvieron éxito en el pasado. En las sociedades ágrafas, la información sobre tales comportamientos está codificada a menudo como ritual, para asegurar su trasmisión a futuras generaciones. Un ejemplo oportuno está encerrado en el estudio de una aldea de las tierras altas peruanas —aldea moderna, pero prácticamente analfabeta— realizado por mi esposa, Billie Jean Isbell, de la Universidad de Cornell.

Su investigación ha revelado que estas familias campesinas basan su independencia económica en un sistema de cultivar una variedad de plantaciones a diferentes alturas. Los aldeanos tienen un nombre específico para cada zona de plantación. La diferencia entre las zonas ha quedado simbolizada mediante la construcción de una capilla en el límite de cada una. El ritual anual de la cosecha incluye una serie de visitas a las capillas y la recolección de una muestra de la producción cultivada en cada zona. De cada ermita se toma una cruz, que es decorada con las plantas cose-

chadas. Todas las cruces se llevan a la aldea, donde son ofrecidas al cura de la misma.

Aun en el caso de que un joven aldeano dejase de aprender de su familia o de los viejos de la aldea el oficio de campesino, la lección de este ritual anual difícilmente podría pasarle inadvertida: "Aquí están las partes que dividen la tierra. Aquí están las cosechas que se producen en cada parte. Agrada a Dios (y al sacerdote) que todas estas cosechas se planten cada año". Una comprensión consciente de las ventajas económicas de los cultivos mixtos puede que ni siquiera exista entre los aldeanos, pero el esquema de comportamiento que el ritual refuerza simbólicamente contribuye a proteger a sus practicantes del peligro de fallo de la cosecha única y del trastorno social que acarrea la dependencia del trabajo a jornal.

R. Tom Zuidema, de la Universidad de Illinois, ha logrado reconstruir rituales similares, observados por el pueblo de Cuzco cuando esta ciudad era la capital del imperio de los incas. En aquella época, líneas imaginarias irradiaban en todas direcciones desde Cuzco. Eran líneas imaginarias en tanto que no estaban señaladas, pero la orientación de cada una se hallaba indicada por una serie de santuarios. Cada día del año un grupo afín diferente, entre los pobladores de la ciudad, adoraba en un santuario también diferente: de hecho el suelo del valle de Cuzco formaba el mapa de un calendario ritual anual. Por este medio, se comunicaba al pueblo de Cuzco la pertinente información sobre el ciclo agrícola, obligaciones sociales, actividades militares y otras muchas cuestiones. El hecho de que los incas se preocupaban de esta clase de cartografía informativa lo confirma el plano de la capital misma. Los incas llamaban a la ciudad de Cuzco "el puma" y a sus habitantes "miembros del cuerpo del puma". La ciudad fue trazada en forma de puma, aunque la figura quedó algo descompuesta por la necesidad de adaptarla a la topografía del valle.

He sugerido que los geoglifos de Nazca fueron en primer término el producto de mecanismos sociales destinados a regular el equilibrio entre recursos y población. Los descubrimientos de Zuidema sobre la práctica incaica de mapas simbólicos en el terreno (incas que en último término asumieron la autoridad en la región de Nazca) y la documentación reunida por Billie Jean Isbell sobre el ritual agrícola entre los indios de las tierras altas del Perú actual indican que los geoglifos de Nazca contenían también cierta información simbólica, cartografiada en el suelo para que sucesivas generaciones la observaran, reconocieran y memorizaran.

¿Qué eran estos mensajes en forma de mapas? Como ha indicado Reiche durante muchos años, ciertas líneas de la Pampa Colorada señalan la posición del sol en los solsticios de verano e invierno y algunas otras líneas parecen tener también significados propios de calendarios. Un análisis por ordenador realizado por Hawkins sobre la orientación de las líneas, aunque no llegó a demostrar que la mayoría de ellas tuvieran significado astronómico, puso de manifiesto que las orientadas en relación a posiciones extremas anuales del sol y de la luna eran hasta el doble de lo que cabría esperar en base del azar.

Ambos estudios indican que por lo menos algunos de los geoglifos de Nazca tienen posibilidades de relación con el calendario. El mapa de datos de calendario en el terreno, si estaba combinado con observaciones rituales, no sólo podía trasmitir información de significado agrícola sino también informaciones de otras clases, útiles para una sociedad compleja, aunque preliteraria, afrontada con una creciente necesidad de archivar el conocimiento adquirido por la experiencia de generaciones.

Volviendo a las preguntas que formulé al principio: ¿Pueden la arqueología y disciplinas afines determinar cuándo se realizaron los geoglifos de la Pampa Colorada, quién los hizo y por qué? Parece claro que las respuestas a las dos primeras preguntas son que la mayoría, si no todos los dibujos, son obra del mismo pueblo que modeló y pintó la linda cerámica de Nazca entre el 200 a.C. y el 600 d.C. Las tumbas de este pueblo campesino prehistórico y las ciudades y aldeas arruinadas donde un día habitó se encuentran cerca de los geoglifos.

En lo que se refiere a la cuestión del por qué, parece claro también que la factura de los dibujos, como la construcción de los enormes edificios de adobe en el norte, servía para regular los aumentos de población en relación con los cambios de la energía disponible. Naturalmente, una cosa es reconocer la función para que sirven las inversiones de energía de esta clase reguladora y otra muy distinta llegar a comprender la forma tangible asumida por esa inversión. Yo, sin embargo, considero como hipótesis fiable la de que los geoglifos reflejan la necesidad general de las varias sociedades preliterarias del Perú de registrar, o quizá más exactamente, de almacenar la información significativa de cómo funcionaba su sistema.

El almacenaje de inventarios por medio del conocido sistema peruano de enumeración a base de cuerdas y nudos, el llamado quipú, fue evidentemente una práctica adoptada temprano en la prehistoria de los Andes. Datos de calendario, en particular cuando su empleo exige una comprobación de observaciones astrónomicas efectivas, pudieran ser imprácticos o imposibles de archivar por medio de quipús. Yo sugiero que las informaciones de este género fuesen codificadas simbólicamente y registradas en el más duradero de los medios disponibles: la propia superficie de la tierra.

Si los geoglifos de la Pampa Colorada fuesen los únicos del Perú, tal hipótesis tendría poco peso. Pero los geoglifos han aparecido en otros muchos lugares. Una figura rectangular de 60 metros de largo y 30 de ancho, que encierra "decoraciones" geométricas, ha sido registrado en el mapa de una zona por encima del valle de Sihuas, al sur de Nazca. Líneas de "pistas" están presentes en el valle de Ica, un poco al norte de Nazca; en las cercanías de Lima, aún más al norte; al norte de Lima, en el valle de Virú, cerca de la monumental Huaca del Sol; y según informaciones, en el valle de Zaña, a 180 kilómetros al norte de Virú. Alberto Carbajal, Carlos Williams y yo fotografiamos recientemente líneas no registradas con anterioridad, figuras geométricas, cadenas de espirales y un geoglifo de un rostro humano de 43 metros de ancho. Fotografías aéreas tomadas a poca altura y exploraciones de otras zonas del suelo de los desiertos peruanos podrían descubrir muchas más figuras. Es evidente que se empleó energía para producir geoglifos incluso en partes del Perú donde la mayor parte de la inversión de trabajo comunal tuvo por objeto la construcción de plataformas de adobe y otras obras monumentales. Los geoglifos que bordean la costa septentrional y central pueden ser modestos comparados con los de Nazca, pero demuestran que tales figuras no eran exclusivamente un fenómeno de la costa sur.

En los siglos que siguieran al Período Intermedio Temprano la energía invertida en plataformas de templos continuó produciéndose, pero a una escala más pequeña y menos impresionante. Lo mismo parece haber ocurrido con los geoglifos. Evidentemente la naciente autoridad estatal centrada en Huari tomó medidas para que el principal esfuerzo comunal se desviase hacia la construcción de edificios administrativos, medios de almacenamiento, fortificaciones y lo que parecen haber sido centros de manufacturas. La autonomía regional dio paso a la autoridad centralizada, preparando la escena para el nacimiento final del imperio inca. Podemos esperar en buena lógica que las antiguas prácticas de archivos comarcales, de datos y búsquedas de los mismos, escritos en la tierra y envueltos en el ritual local, fuesen dejando poco a poco de existir.

La formación de modelos en el desarrollo biológico

En el embrión de muchos organismos, las células en proceso de diferenciación se organizan en el espacio enterándose primero de dónde están exactamente e interpretando luego esta información de acuerdo con sus programas genéticos

Lewis Wolpert

¿Cómo se da el desarrollo de una sola célula, el óvulo fertilizado, dando lugar a un intrincado sistema de muchos tipos celulares diferentes, organizados según modelos predecibles, para formar un tipo determinado de animal adulto? Para contestar a esta pregunta central de la biología del desarrollo, la mayoría polarizan su atención en cómo las células del huevo en división llegan a diferenciarse unas de otras, originando los diversos tipos celulares, desde el cartílago y el músculo hasta las células nerviosas y hemáticas, de que consta un animal adulto. Puesto que cada tipo celular se caracteriza por la presencia de proteínas particulares, y dado que se acepta de manera general que la información genética en cada célula es la misma y permanece inalterada durante el desarrollo, el interés se ha puesto, principalmente, en averiguar cómo se regulan los genes que codifican las proteínas específicas, cómo se activan o desactivan en los diferentes tipos celulares.

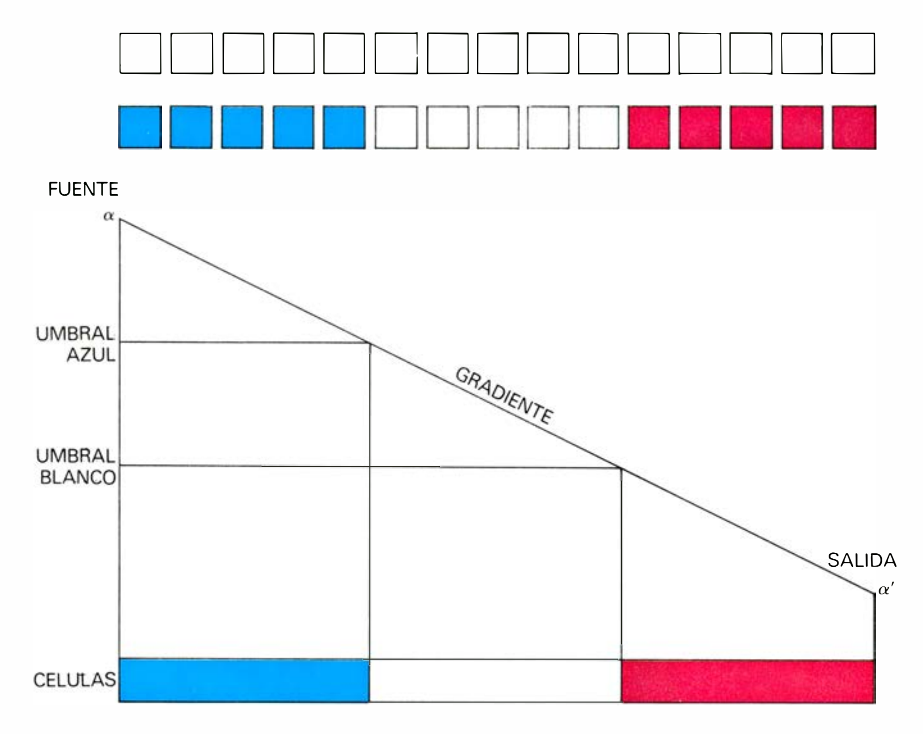
Pero la diferenciación celular es sólo un aspecto del desarrollo. Las diversas formas que caracterizan al animal adulto surgen en gran parte de un proceso bastante diferente: la formación de modelos. Consideremos el brazo y la pierna. Ambos contienen los mismos tipos de células diferenciadas –músculo, cartílago, tejido conjuntivo y piel–, pero la organización espacial de las células es distinta. Esta organización espacial de tipos celulares diferenciados constituye la esencia de la formación de modelos. El histólogo reconoce unos 200 tipos de células en el cuerpo humano, y el chimpancé tiene exactamente los mismos tipos celulares; la diferencia entre un hombre y un chimpancé ocurre en la organización espacial de las células. Entre los vertebrados, considerados como un todo –peces, anfibios, reptiles y mamíferos–, hay una cierta variación en el tipo celular, pero la clave de la diferente organización de todas estas formas no radica en las células como tales; radica en cómo se orientan estas unidades básicas de construcción en el espacio durante el desarrollo. Insistamos: saber cómo se diferencian el cartílago, el músculo y el tendón, dice muy poco acerca de cómo adquieren su forma los cinco dedos de la mano.

Para entender la formación de modelos nos vamos a valer de una analogía con la bandera nacional francesa. Consideremos una agrupación lineal de células, cada una de las cuales pueda diferenciarse en una célula azul, roja o blanca [*véase la ilustración superior de la página 80*]. ¿Cómo pueden organizarse las células en la fila para formar una bandera francesa, un tercio azul, un tercio blanca y un tercio roja? Hay un cierto número de soluciones posibles del problema, pero la obvia es que las células lleguen a "saber" su posición con respecto a los extremos de la fila y que se programen para diferenciarse adecuadamente a partir de dicho conocimiento sobre la posición. Si una célula está en el tercio derecho de la fila, se diferenciará como roja; si en el tercio central, como blanca, si se halla en el tercio izquierdo, como azul. Un mecanismo de este tipo para la formación de modelos se puede considerar como un proceso de dos pasos: primero, se asigna a las células la información de posición y, luego, ellas interpretan esa información de acuerdo con su programa genético.

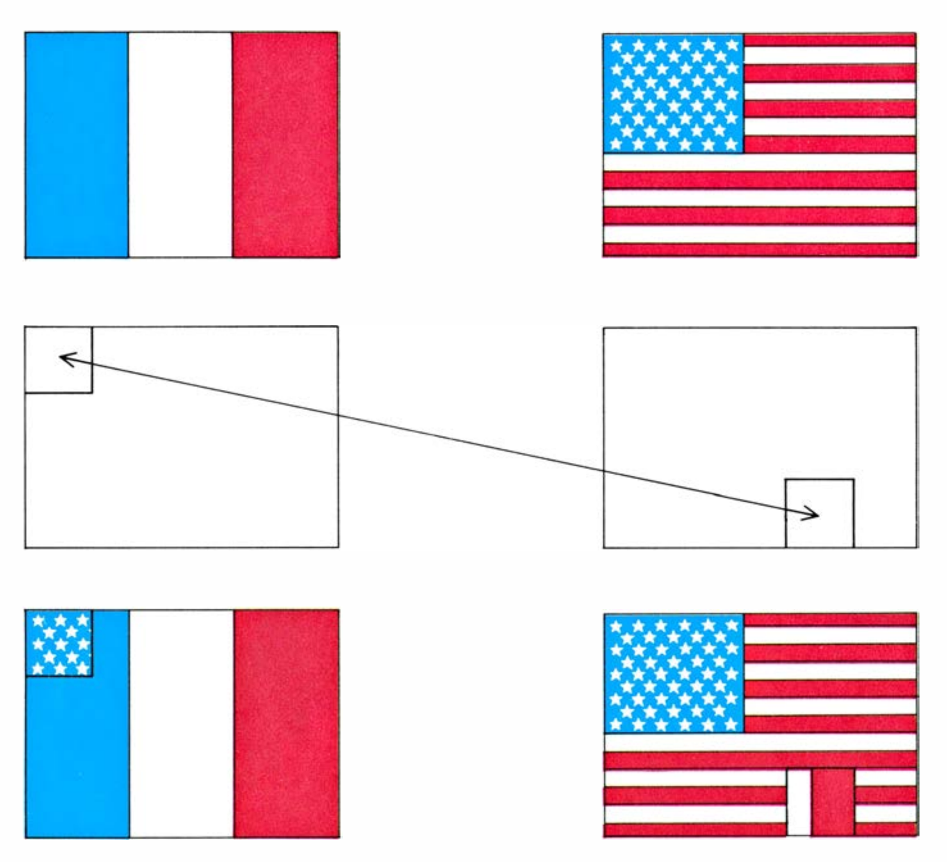
Este mecanismo, muy sencillo en su concepción, tiene algunas características interesantes. La única comunicación que las células necesitan intercambiarse es el mensaje que especifica la información de posición, que puede ser muy elemental. Además, las mismas interacciones y la información de posición pueden dar lugar a varios modelos según el programa de desarrollo específico de acuerdo con el cual se interpreta la información. Por ejemplo, la bandera nacional estadounidense consta de células de los mismos colores que la bandera francesa. Si los dos emblemas se desarrollan a partir de la información de posición especificada por el mismo sistema coordenado bidimensional, entonces un injerto de un trozo de bandera en la otra debería comportarse de acuerdo con su posición y su programa genético. Una pequeña porción de bandera estadounidense "indiferenciada", injertada en el borde superior izquierdo de la bandera francesa, debería desarrollarse dando estrellas blancas sobre un campo azul; injertada cerca del ángulo inferior derecho, debería desarrollarse y formar bandas horizontales rojas y blancas.

En los sistemas biológicos hay buena muestra de este tipo de desarrollo. Curt Stern, de la Universidad de California en Berkeley, y sus colaboradores consiguieron, al construir mosaicos genéticos, lo que son efectivamente estos injertos. En la mosca del vinagre *Drosophila*, una mutación en un solo gen da lugar a una forma mutante, *aristapedia*, con una pata donde debería estar la antena. Mediante manipulación genética se pueden obtener tanto células tipo salvaje, o células normales, como células *aristapedia* en la región antenal. Las células se comportan de acuerdo con su posición y sus diferentes programas genéticos. Si las células del extremo de la antena son del tipo salvaje, forman las estructuras normales del extremo de la antena, pero si portan el gen *aristapedia* forman estructuras de la pata, pero no cualesquiera de ellas sino las estructuras del extremo de la pata. Esto sugiere que la información de posición (¿a qué distancia del extremo?) es la misma en la antena que en la pata; lo que difiere es la interpretación. Otros mutantes del modelo se comportan de manera similar en los mosaicos, desarrollándose las células de acuerdo con su posición y su constitución genética. Existe, pues la intrigante

ESBOZO DEL ALA DE UN POLLUELO, en sucesivos estadios de su incubación: después de cuatro, cinco y seis días (*arriba*), siete y ocho días (*en el centro*) y 10 días (*abajo*). Las estructuras cartilaginosas (*véase la ilustración de la parte superior de la página 82*) son fácilmente reconocibles en una preparación entera teñida del tejido, ya que capta un colorante específico. Las fotografías fueron hechas por Dennis Summerbell y Jim Smith en el laboratorio del autor.



ANALOGIA DE LA BANDERA, que muestra cómo una fila de células (*arriba*) que pueden diferenciarse en células azules, blancas o rojas (*segunda fila*) pueden organizarse para formar la bandera francesa. Aquí, la información de posición viene proporcionada por un gradiente de morfógeno difusible; las células tienen umbrales adecuados. Una fuente mantiene el gradiente de concentración fijo en α en un extremo y una salida mantiene la concentración en α' en el otro. Los umbrales son propiedades que permiten interpretar el gradiente en información de posición. Así, por encima de la concentración U_A una célula se convierte en azul.



DESARROLLO DEL INJERTO en la bandera, que ocurre según la posición y el "programa genético". El material indiferenciado de la bandera estadounidense injertado en la esquina superior izquierda de la bandera nacional francesa se desarrolla bajo la forma de estrellas; de manera similar, una porción de la bandera francesa injertada en la bandera nacional de los Estados Unidos se desarrolla de acuerdo con su programa genético, el cual, cerca de la parte inferior corresponde a bandas verticales blancas y rojas.

posibilidad de que la información de posición sea universal.

La idea de que la posición determina cómo van a desarrollarse las células fue establecida explícitamente por Hans A. E. Driesch a finales del siglo XIX; estaba implícita en las teorías del gradiente sobre la formación de modelos defendidas en la década de 1920 por C. M. Child, de la Universidad de Chicago. La consideración de la formación de modelos en términos de información de posición permite analizar, con rigor, este aspecto del desarrollo y obliga a pensar acerca de los mecanismos a través de los cuales se especifica la posición. La determinación de la posición implica un sistema de coordenadas; y un sistema de este tipo, a su vez, requiere una "región frontera" especial, con referencia a la cual se mide la posición; así pues: polaridad, que determina la dirección en que se mide la posición, y una cierta cantidad, para medir la posición. Una interpretación basada en la información de posición hace más sólido el viejo concepto de campos embrionarios, dentro de los cuales ocurre y se regula el desarrollo: los campos pueden ahora considerarse como sistemas coordenados especiales. Si la posición se determina por un sistema coordenado, entonces la información de posición puede resultar parecida en organismos muy diferentes.

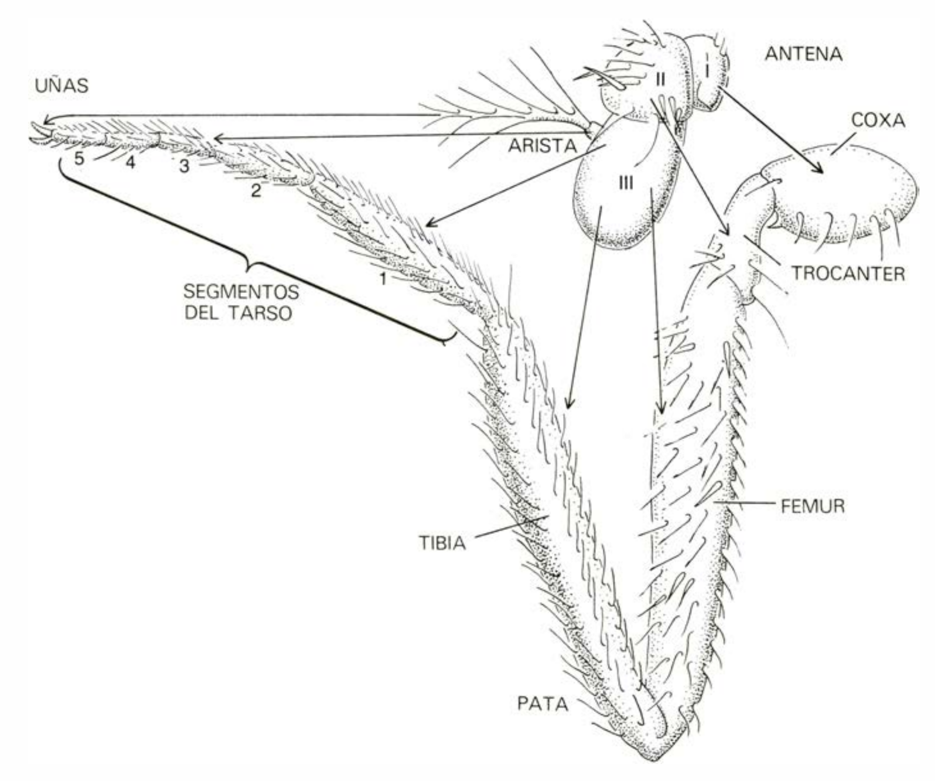
¿Cómo puede determinarse la posición? Podría ser una manera establecer un gradiente, mediante alguna propiedad que disminuyera de forma continua al aumentar la distancia, en un campo de posición, a una región frontera o de referencia. Una característica general de los campos de posición es que siempre son pequeños, y, otro rasgo, que el tiempo exigido para establecerlos dura mucho. Todos los campos de posición conocidos no llegan a albergar 100 células (cerca de un milímetro de longitud), y por lo común resultan bastante más pequeños; el tiempo requerido para formarlos es del orden de horas. Observando estos hechos, Francis H. C. Crick señaló en 1970 que una manera plausible de determinar la posición sería la difusión de un "morfógeno" a través del campo, una substancia cuya concentración pudiera percibirse por células individuales. Dada una fuente de una determinada concentración del morfógeno en un extremo del campo y una salida en el otro extremo, podría establecerse un gradiente lineal a través del campo. Más tarde Alfred Gierer y Hans Meinhardt, del Instituto Max Planck de Investigación Vírica de Tubinga, demostraron cómo puede establecerse un gradiente de difusión sin una fuente o salida localizadas.

Sin importar cómo se establezca un gradiente de este tipo, las células pueden

interpretarlo si su programa genético está especificado en términos de umbrales; es decir, si por encima de cierta concentración las células se diferencian según un tipo y por debajo de ella en otro. En mi laboratorio de la Middlesex Hospital Medical School, de Londres, Julian Lewis, Jonathan Slack y yo mismo hemos encontrado que debería ser posible determinar unos 30 estadios celulares diferentes a lo largo de una fila de unas 100 células. El factor que limita la precisión con que esto puede hacerse es la exactitud con que pueden establecerse los umbrales para células individuales, cosa que a su vez depende de la exactitud con que las células pueden controlar la concentración de las moléculas en su interior. No se conoce esa propiedad celular, que se diría fundamental, pero parece razonable una exactitud con un 20 por ciento de error. Parte de la variabilidad característica de los procesos del desarrollo (y de otros procesos celulares) debe surgir de esta variabilidad intrínseca.

Dennis Summerbell, Lewis y yo hemos propuesto otro modelo, específicamente para sistemas en crecimiento, en el cual las células calculan la distancia midiendo el tiempo que permanecen en una región particular que llamamos "zona de progreso". Al hallarse las células en proceso de división y permanecer constante la amplitud de la zona, las células están siempre abandonando la zona: quedan atrás al progresar la zona. Las células que salen en primer lugar están al principio de la fila (en el sentido de la fila de células en la analogía de la bandera) y las células que salen en último lugar estén al final de la fila; si las células pueden medir el tiempo que permanecen en la zona de progreso, habrán adquirido información respecto a su posición.

Ilustraré, en primer lugar, los modelos de gradiente de difusión y zonas de progreso con respecto al desarrollo del ala del polluelo. Hago notar que estos modelos parecen implicar interacciones de una célula a otra bastante sencillas. Está muy extendida la opinión de que las células sostienen complejas conversaciones una con otra a lo largo del desarrollo. Habrá que ir pensando en la posibilidad de que, por lo menos en la formación de modelos, sus conversaciones sean, por el contrario, muy limitadas e incluso aburridas.

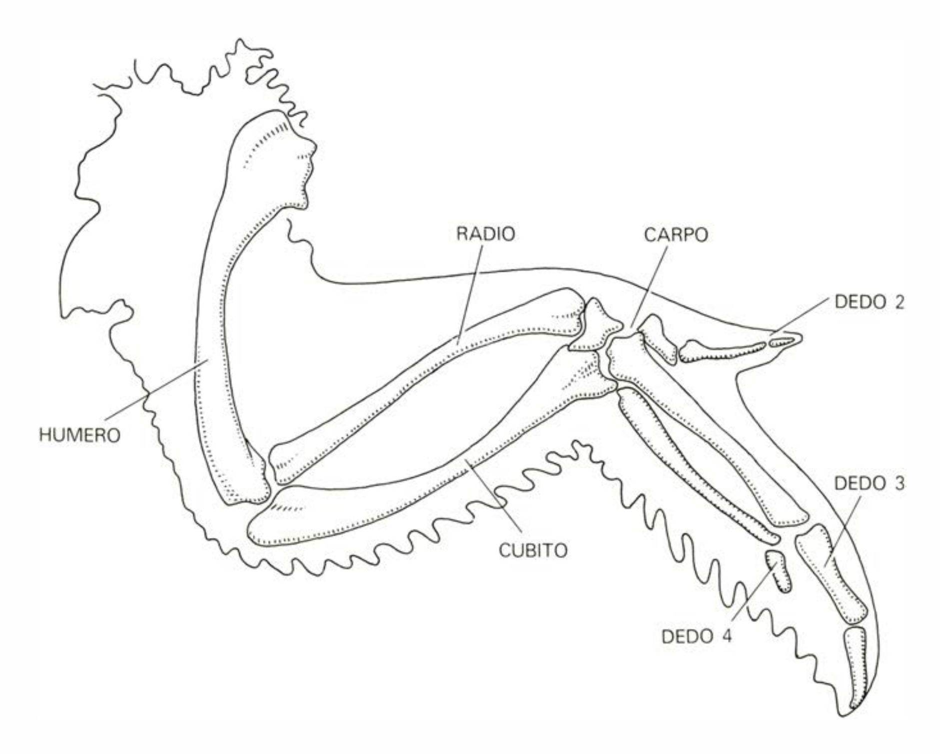


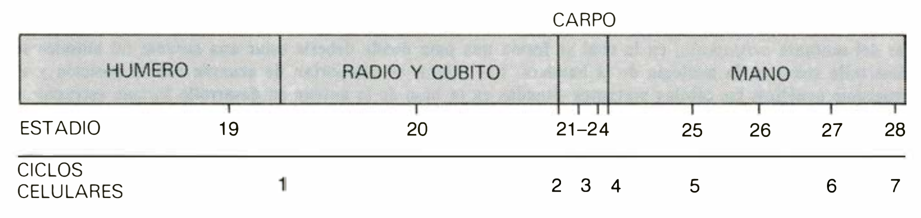
MOSAICO GENETICO que combina las células de la mosca del vinagre normal *Drosophila melanogaster* y las del mutante *aristapedia*, en la cual se forma una pata donde debería estar una antena; tal mosaico se desarrolla como en la analogía de la bandera. Las células se comportan de acuerdo con la posición y el programa genético: las células mutantes situadas en la base de la antena en desarrollo forman estructuras basales de la pata en desarrollo, mientras que las que están cerca del ápice de la antena forman las uñas o segmentos del tarso. Las flechas indican las regiones de la pata y de la antena que parecen tener valores de posición similares. La idea de que la posición determina cómo van a desarrollarse las células fue establecida por Hans A. E. Driesch. Este dibujo es una adaptación del trabajo de Howard A. Schneiderman y John Postlethwait, investigadores que se encuentran actualmente en la Universidad de California en Irvine.

El ala de un polluelo se desarrolla a partir de un esbozo en forma de paleta muy parecido al esbozo del brazo de un embrión humano. Nace como una pequeña abolladura, que aparece a los dos días de la puesta del huevo y en un momento en que las estructuras axiales principales, como los somitos, ya se han formado. El modelo de cartílago, músculo y tendones toma forma dentro de una holgada red de células mesenquimatosas encerradas dentro de una lámina de células ectodérmicas, que se convertirán en piel. Después de 10 días en la incubadora, el modelo básico de los huesos de la extremidad -húmero, radio y cúbito, carpo y dedos- está bien establecido como cartílago (gran parte del cual se convertirá más tarde en hueso). En preparaciones fijadas, el cartílago se tiñe con un colorante que lo hace visible con nitidez, a través de la piel.

La mayoría de los descubrimientos importantes sobre las interacciones dentro del esbozo en desarrollo se deben a John W. Saunders, Jr., de la Universidad estatal de Nueva York en Albany, y sus colegas. Encontró que una región engrosada del ectodermo, que cruza a modo de reborde el extremo del esbozo del miembro, era esencial para el desarrollo. Si se extirpa este reborde apical ectodérmico mientras se desarrollan los elementos cartilaginosos, ciertas partes de la extremidad suspenderán su formación. Cuanto más tarde se extirpe el reborde, más se desarrollarán las estructuras. Si el reborde se elimina en una fase precoz, sólo llegará a formarse, posiblemente, el húmero; si se extirpa más tarde, quizá sólo dejen de formarse los dedos. Esto muestra que las estructuras están determinadas según una secuencia proximodistal, es decir, empezando cerca del punto de implantación de la extremidad en el cuerpo y siguiendo en dirección a la parte distal del miembro. Esta secuencia se observa claramente en el esbozo que se está gestando, en el cual el húmero se hace visible mucho antes de que aparezcan los dedos.

Si el modelo de la diferenciación celular en el esbozo de la extremidad avanza basándose en la información de posición, se puede considerar que las células tienen asignados unos valores de posición en un sistema coordenado tridimensional; la interpretación de estos valores de posición daría lugar al modelo de diferenciación observado para el cartílago, músculo y tendón. Sería algo así como esculpir mediante números. Los tres ejes naturales de la extremidad son el eje proximodistal (de la parte superior de la extremidad hacia los dedos), el anteroposterior (de delante hacia atrás) y el dorsoventral (de arriba abajo). Aquí consideraré únicamente los dos primeros de estos ejes, para los cuales Summerbell, Lewis, Cheryll Tickle y yo hemos propuesto dos mecanismos diferentes en orden a determinar la información de posición. Para el eje proximodistal proponemos un mecanismo en el que el tiempo proporciona información respecto a la posición. El mecanismo que sugerimos para el eje anteroposterior es una señal procedente de una región frontera, que pone un gradiente de valores de posición.





ELEMENTOS CARTILAGINOSOS de la extremidad del polluelo. La barra que subyace al dibujo muestra el efecto sobre el desarrollo de la extremidad cuando el reborde ectodérmico apical, un engrosamiento a modo de paleta en el extremo del esbozo de la extremidad, se extirpa en momentos diferentes. El tiempo se indica en términos de una serie estándar de estadios del desarrollo y en términos de ciclos celulares. Por ejemplo, si el reborde se extirpa en el estadio 20, sólo crecerán el húmero y la mitad del radio y cúbito. Aquí cabe destacar que, para dar lugar a la corta (aunque compleja) región del carpo, son necesarios dos ciclos celulares.

La determinación de los valores de posición proximodistales parece tener lugar dentro de una zona de progreso cercana al extremo terminal del esbozo del miembro. La zona está siempre asociada con el reborde apical ectodérmico, y alguna influencia procedente del reborde es la que presumiblemente condiciona su aparición. La zona de progreso es de unos 350 micrometros (milésimas de milímetro) de largo. Contiene una población de células en división que, por su aspecto general, recuerdan a los fibroblastos. Supongamos, por convención, que el valor de posición disminuye autónomamemte con el tiempo de permanencia en la zona, esto es, cuanto más tiempo una célula permanezca en la zona, más bajo será su valor de posición.

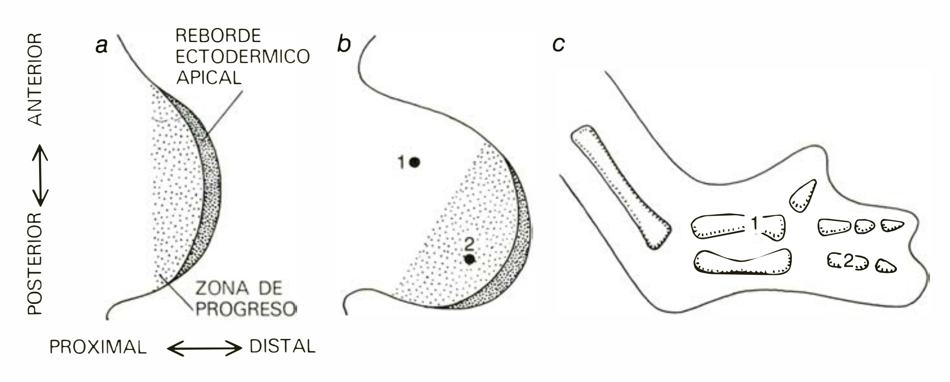


DIAGRAMA DEL ESBOZO DE LA EXTREMIDAD visto desde arriba, mostrando el eje proximodistal (desde el lugar de unión hacia el extremo) y el eje anteroposterior (de delante hacia atrás). De acuerdo con el modelo del autor, los valores de posición a lo largo del eje proximodistal están determinados por el tiempo que una célula pasa en la zona de progreso, que está asociada con el reborde ectodérmico apical (*a*). Para las células de la región 1 su posición ha sido ya determinada; la posición de las células en la región 2 se halla en fase de determinación (*b*); por último, las dos regiones acabarán formando estructuras diferentes (*c*).

Al hallarse todas las células en proceso de división, estarán continuamente saliendo de la zona, y las células que la abandonen primero formarán estructuras proximales, mientras que las que lo hagan tardíamente darán lugar a las puntas de los dedos; el gradiente de valor de posición generado por este proceso determinará la serie de elementos a lo largo del eje proximodistal. Ahora puede verse por qué la eliminación del reborde ectodérmico apical da lugar al truncamiento de la extremidad: la zona de progreso queda abolida, de manera que no se generan nuevos valores de posición.

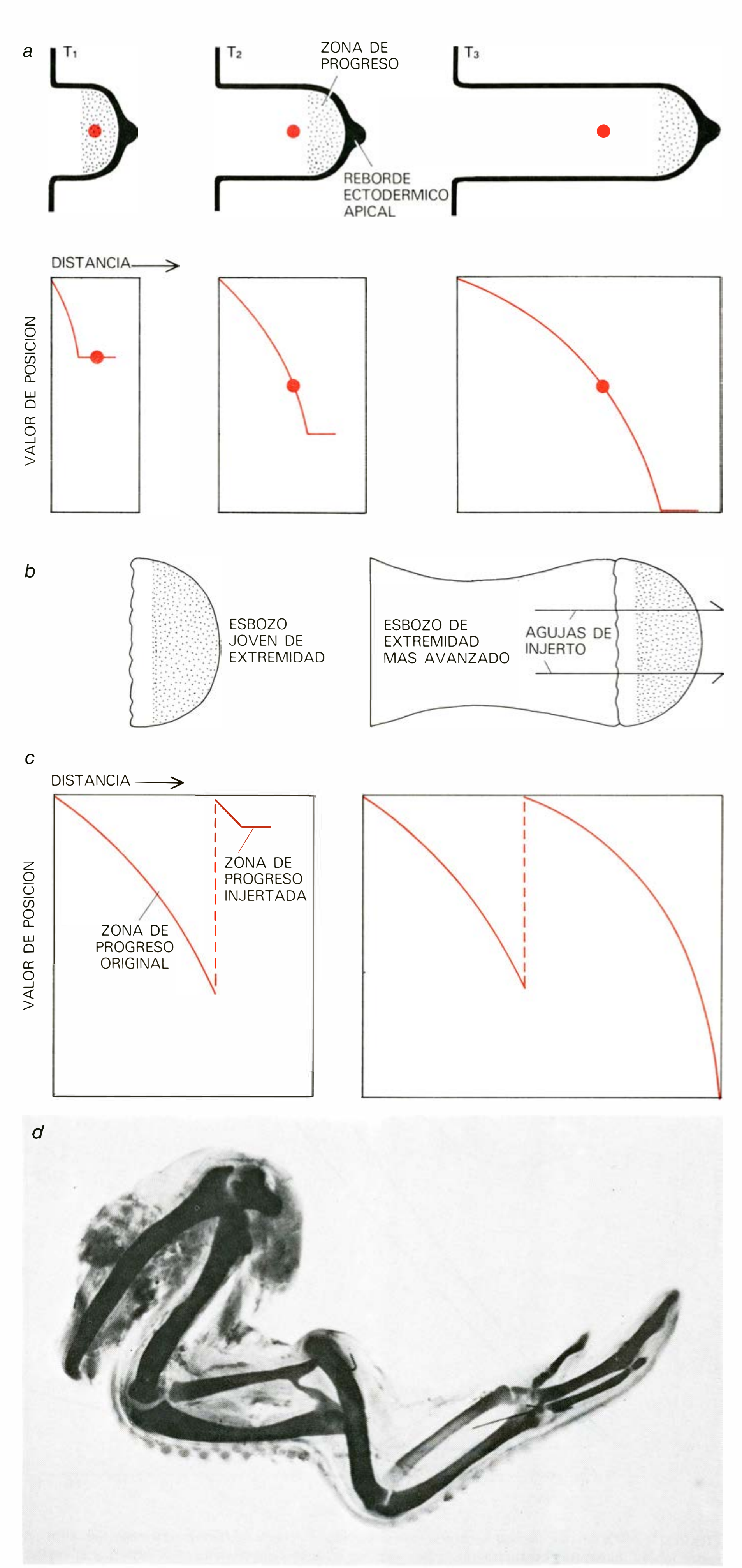
Este modelo sugiere que si no hay interacciones a larga distancia entre las células mesenquimatosas, una zona de progreso continuaría desarrollándose autónomamente cuando se extirpara e injertara en otro sitio. Si la zona de un esbozo joven se injertara sustituyendo la zona equivalente en un esbozo en el que los valores de posición que corresponden al húmero y al radio y cúbito han sido ya generados, la zona injertada debería generar un segundo conjunto de tales valores; esto daría lugar a la duplicación de elementos a lo largo del eje proximodistal. El experimento inverso, en el que la zona de progreso más vieja se injerta en un esbozo joven, daría lugar a la ausencia de estructuras proximales: aparecerían los dedos, pero faltaría el húmero, el cúbito o el radio. Nuestros injertos experimentales concuerdan bastante bien con la teoría [*véase la ilustración de la página opuesta*]. (Debo señalar, sin embargo, que Madeleine Kieny y sus colegas de la Universidad de Grenoble han obtenido resultados diferentes, encontrando menos autonomía de este tipo y más pruebas de regulación e interacción encaminadas a conseguir un juego de estructuras más normal.)

No nos fue fácil conseguir posteriores pruebas experimentales para el modelo de la zona de progreso, pero algunos experimentos recientes de irradiación mediante rayos X han resultado alentadores. Tickle y yo estudiamos el efecto de grandes dosis de rayos X (hasta 2000 rads) sobre el desarrollo del esbozo de la extremidad. Una dosis de este tipo mata el embrión, y, por tanto, injertamos el esbozo irradiado en un sitio de un embrión huésped no irradiado. Encontramos que al aumentar la radiación eran las estructuras proximales las que se perdían; las distales (los dedos) eran relativamente normales. Puede parecer sorprendente que la irradiación afecte menos a las estructuras formadas en una etapa posterior, pero tal es lo que nuestro modelo predice. Si la irradiación mata un gran porcentaje de células en la zona de progreso, las restantes deben permanecer en la zona más tiempo, mientras dicha región se repuebla; adquieren un valor de

posición más bajo, originando así, estructuras distales. De esta forma, el modelo empieza a proporcionar pistas sobre ciertas malformaciones congénitas de las extremidades humanas como la focomelia, o extremidad de foca, que da lugar a una disposición similar de las partes. La droga talidomida podía causar focomelia, entre otras anormalidades, en un feto cuando se administraba a una mujer gestante. Se desconoce su modo de acción, pero nuestros resultados dan por supuesto que de alguna manera podía haber destruido las células en el esbozo precoz de la extremidad del embrión.

Cuando se registra (en términos de ciclos de división celular) el tiempo que las células destinadas a formar unas estructuras particulares permanecen en la zona de progreso, llama la atención el tiempo relativamente largo necesario para formar la muy corta región del carpo. Un análisis llevado a cabo por Lewis y Summerbell justificó la aparente desproporción, que resulta tener importantes implicaciones para el proceso de crecimiento. Los diversos segmentos de la extremidad no difieren mucho en cuanto al tamaño cuando emergen de la zona de progreso; de hecho, su longitud inicial es aproximadamente igual a la cantidad de material que abandona la zona durante un ciclo celular, o unos 350 micrometros. Aunque el carpo es corto, encierra bastante complejidad, estando constituido aparentemente por dos elementos cartilaginosos formados en secuencia, lo cual explica el tiempo necesario para su desarrollo. Es corto, porque apenas crece después de salir de la zona, mientras que sí se desarrollan las otras estructuras de la extremidad. El carpo y el cúbito adyacente empiezan teniendo aproximadamente la misma longitud, pero a los 10 días el cúbito ha prolongado unas 16 veces su longitud inicial, en tanto que el carpo sólo ha doblado su tamaño.

Nigel Holder encontró que esta relativa falta de crecimiento es una propiedad autónoma de la región del carpo, pues persiste cuando el carpo aislado se cultiva o se injerta en otras partes de la extremi-



MEDIDA DEL TIEMPO en la zona de progreso (*área punteada*); se demuestra para un grupo de células (*punto coloreado en el dibujo a*) después de uno, dos y tres ciclos celulares (T_1-T_3). El valor de posición de las células decrece (*curva coloreada*) mientras permanecen en la zona, pero se mantienen constante después de haber abandonado la zona. Si la zona de progreso de un esbozo joven de extremidad se injerta sustituyendo la zona de un esbozo más avanzado (*b*), y si las zonas de progreso se comportan autónomamente, debería generarse un nuevo gradiente en tándem con el gradiente original de valor de posición (*c*). Esto puede dar lugar a que se repitan las estructuras a lo largo del eje proximodistal. Una posible secuencia sería húmero, radio y cúbito, húmero, radio y cúbito, carpo y dedos. Esta secuencia se muestra en la fotografía de Summerbell de la extremidad (*d*).

dad. Parece ser que los valores de posición impartidos al cartílago dan lugar a diferentes programas de crecimiento en diferentes regiones; dos trozos de cartílago que parezcan similares pero que estén en distintas regiones pueden mostrar propiedades diferentes. Lewis y yo hemos llamado "no equivalencia" a esta situación en la cual células de la misma clase de diferenciación tienen distintas propiedades programadas por distintos valores de posición. La longitud final de los elementos cartilaginosos está determinada por su tamaño al salir de la zona de progreso y por su programa de crecimiento; ambos determinantes se especifican por valores de posición. Estos valores pueden alterarse durante la evolución, y se supone que las posibles alteraciones justifican algunas de las diferencias en el tamaño y forma de las extremidades de los vertebrados.

Nuestro modelo de gradiente de morfógeno para el eje anteroposterior se basa en otro descubrimiento de Saunders. Descubrió una pequeña región de células en el borde posterior del esbozo del ala que, si se injertaban en un lugar próximo a la parte anterior del esbozo, podían provocar que la extremidad formara dedos adicionales: algo así como una duplicación a modo de imagen especular a lo largo del eje anteroposterior. Creemos que una señal graduada procedente de esta región, que se llama zona de actividad polarizante, proporciona información de posición a lo largo del eje anteroposterior. Nuestra hipótesis es que se establece un gradiente en un morfógeno que sale de la zona y que los diversos dedos, por ejemplo, están especificados a diferentes umbrales. Los dedos, numerados convencionalmente 2, 3 y 4, son marcadores idóneos porque se distinguen fácilmente uno del otro en el ala del polluelo. Con el fin de comprobar la hipótesis injertamos, por tanto, una zona de actividad polarizante en sucesivos lugares a lo largo del eje anteroposterior y estudiamos si el modelo de los dedos estaba de acuerdo o no con las predicciones del modelo.

Si la información de posición se libera a lo largo de un gradiente y se interpreta en términos de umbrales, entonces el dedo número 4, el más próximo a la zona normal de actividad polarizante, debería ostentar el umbral más alto y el dedo número 2 debería tener el más bajo. Una nueva zona de actividad polarizante injertada bastante adelante en el esbozo, con la zona original dejada en su posición posterior normal, debería originar una duplicación especular [*véase la ilustración de la página opuesta*]. Esto es, el modelo de los dedos sería 4, 3, 2, 2, 3, 4. Y así ocurrió en nuestro experimento. Si se coloca el injerto más cerca de la parte media del esbozo, el gradiente debería ser tal que el modelo previsto (leyendo desde el borde anterior del ala al borde posterior) sería 2, 3, 4, 4, 3, 3, 4; nosotros obtuvimos extremidades con aproximadamente este modelo. En conjunto, los modelos observados en varios experimentos encajaban bastante bien con los modelos previstos.

Tres observaciones posteriores apoyan la idea de una señal graduada. Primero, si se coloca una segunda zona de actividad polarizante al lado de la zona del huésped, no tiene efecto aparente; esto es lo que hay que esperar si la zona actúa como fuente de un morfógeno cuya concentración se regula para permanecer fija. Segundo, si el injerto se coloca muy hacia adelante, no se forman dedos adicionales; la señal no es suficientemente fuerte para subir por encima del umbral. A medida que el injerto se implanta sucesivamente más hacia atrás, el primer dedo adicional que aparece es el número 2; que es el que tiene el umbral más bajo. Tercero, el modelo predice que, al atenuar progresivamente la señal procedente de una zona de actividad polarizante, debería interferir sucesivamente con la formación de los dedos 4, 3 y 2. Jim Smith sometió zonas a irradiación extremadamente alta de rayos X antes de injertarlas en una posición para la cual el modelo resultante de dedos es normalmente 4, 3, 2; encontró que con el aumento de irradiación (y el consiguiente incremento de la atenuación) era el dedo número 4 el que en realidad primero dejaba de aparecer, luego el 3 y finalmente el número 2.

Una de las implicaciones de un mecanismo de formación de modelos basado en información de posición es que las señales como las que tienen un efecto sobre el esbozo de la extremidad del polluelo puedan tener un efecto en todos los vertebrados, con la diferencia de que cada modelo de extremidad refleja diferencias de interpretación. Por ello nos mostramos complacidos cuando Tickle encontró que un trozo de tejido tomado de la supuesta región de la zona de actividad polarizante del embrión de ratón determinó dedos adicionales del polluelo cuando se injertó a un esbozo de extremidad de un polluelo. Jeffrey A. MacCabe, de la Universidad de Tennessee, obtuvo el mismo resultado con embriones de hamster. John F. Fallon, de la Universidad de Wisconsin, ha mostrado que la zona de actividad polarizante de la extremidad de un embrión de tortuga de mar y la de la extremidad de un embrión humano tienen efectos similares en polluelos. La señal parece ser la misma en los vertebrados; lo que ha cambiado en la evolución es la respuesta.

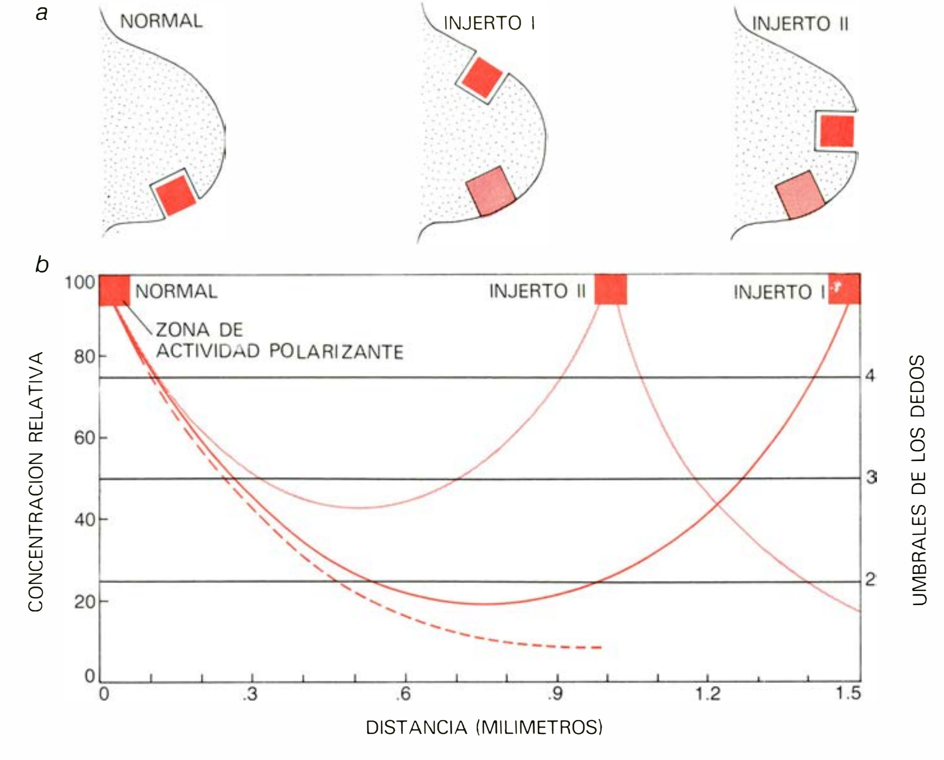
CRECIMIENTO de tres elementos cartilaginosos diferentes; después de haberse formado, los mismos, varían mucho. Summerbell encontró que el carpo apenas crece en comparación con el húmero y el cúbito.

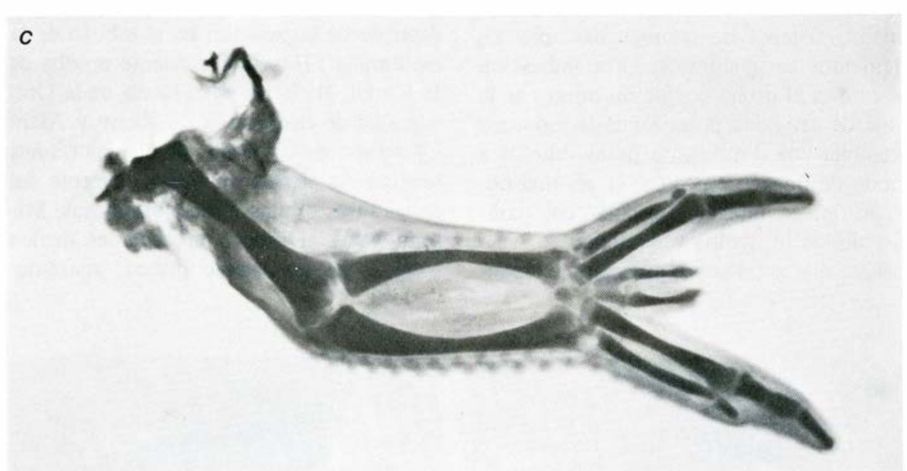
A pesar de la conformidad de los experimentos con nuestro modelo, importa señalar que aún falta una prueba directa para demostrar el morfógeno difusible. Otros modelos, que establecen valores de posición graduados de otras maneras, podrían explicar también los resultados observados.

Hasta el momento sólo he considerado los elementos cartilaginosos de la extremidad del polluelo. El modelo de los músculos y tendones es igualmente importante y algo más complejo. Geoffrey Shellswell y yo hemos examinado el desarrollo del modelo del músculo y tendón, desde nuestra perspectiva sobre la información de posición, concentrándonos en la parte distal de la extremidad: radio y cúbito, carpo y dedos. Aunque los elementos cartilaginosos pueden examinarse fácilmente en la extremidad intacta, para el músculo y el tendón es necesario seguir el tedioso procedimiento de llevar a cabo secciones seriadas, reconstruyendo el modelo a partir de las mismas.

Si el modelo resulta de la interpretación de la información de posición en un sistema coordenado tridimensional -si fuera como esculpir con números- entonces los tendones, músculos y cartílagos deberían desarrollarse autónomamente, sobre la base de posición. Por ello, una de las primeras preguntas que nos hicimos fue: ¿Hasta dónde llega la autonomía, en su desarrollo, del músculo, el tendón y el cartílago? Hay un músculo llamado flexor profundo de los dedos (FPD) que está anclado en el cúbito, cerca del codo, y conectado por un largo tendón al extremo del dedo número 3. Y seguimos preguntando: ¿Depende el desarrollo del tendón del desarrollo de su músculo? ¿Qué pasará con el desarrollo del tendón si falta su lugar de unión normal? Injertamos el extremo de un esbozo precoz de extremidad al flanco del embrión. El tendón del FPD se desarrolló normalmente aunque su músculo no estuviera allí, y el músculo dejado en el muñón del esbozo se desarrolló con normalidad aunque su tendón estuviera ausente.

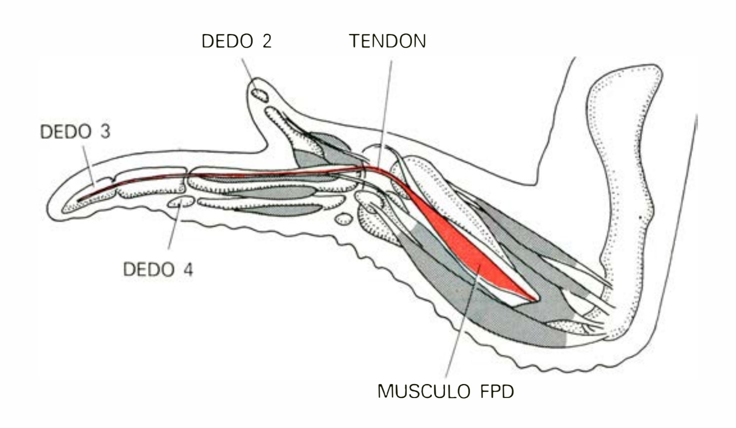
Este y otros experimentos mostraron en general que el músculo, el tendón y el cartílago se desarrollan autónomamente, una vez que su posición haya sido determinada en el esbozo precoz. Parece, sin embargo, existir una regla de busca y únete para los tendones, de modo que rara vez acaban "perdidos" después de una operación de injerto, sino que se unen a tendones, músculos o cartílagos vecinos. Por ejemplo, se puede invertir el extremo del esbozo joven de un miembro dorsoventralmente (lo de arriba irá entonces abajo) mediante el injerto de un extremo de un esbozo de extremidad izquierda a un

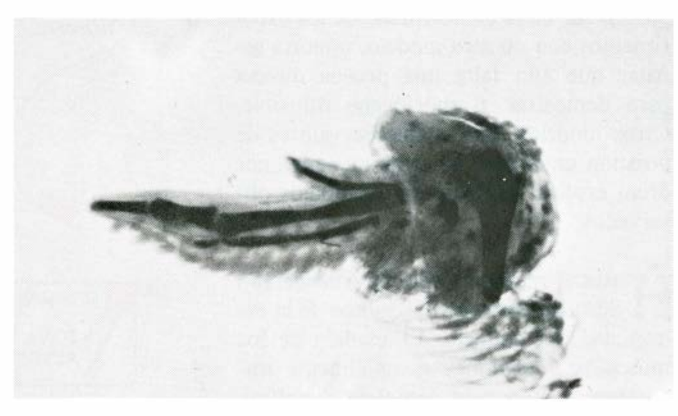






ZONA DE ACTIVIDAD POLARIZANTE (*cuadrado coloreado*); se considera que es la fuente de un morfógeno cuyo gradiente determina la información de posición anteroposterior. Una zona normal en el margen posterior del esbozo (*a*) establece un gradiente (*curva de trazos en* **b**) que interacciona con los umbrales celulares para determinar el modelo normal de los dedos 2, 3 y 4. Si se injerta una zona adicional de actividad polarizante de un esbozo en fase tardía cerca de la parte frontal de un esbozo joven (*I*), el gradiente que crea y la zona huésped de actividad polarizante (*curva ininterumpida en* **b**) determinan una duplicación a modo de imagen especular de los dedos (*fotografía* **c**). Injertando una zona cerca del centro del esbozo (*II*) resultaría un gradiente diferente (*curva fina en* **b**). El resultado experimental se aproximó a lo previsto (*fotografía* **d**). Los injertos fueron realizados por Smith, quien, asimismo, sacó las fotografías.



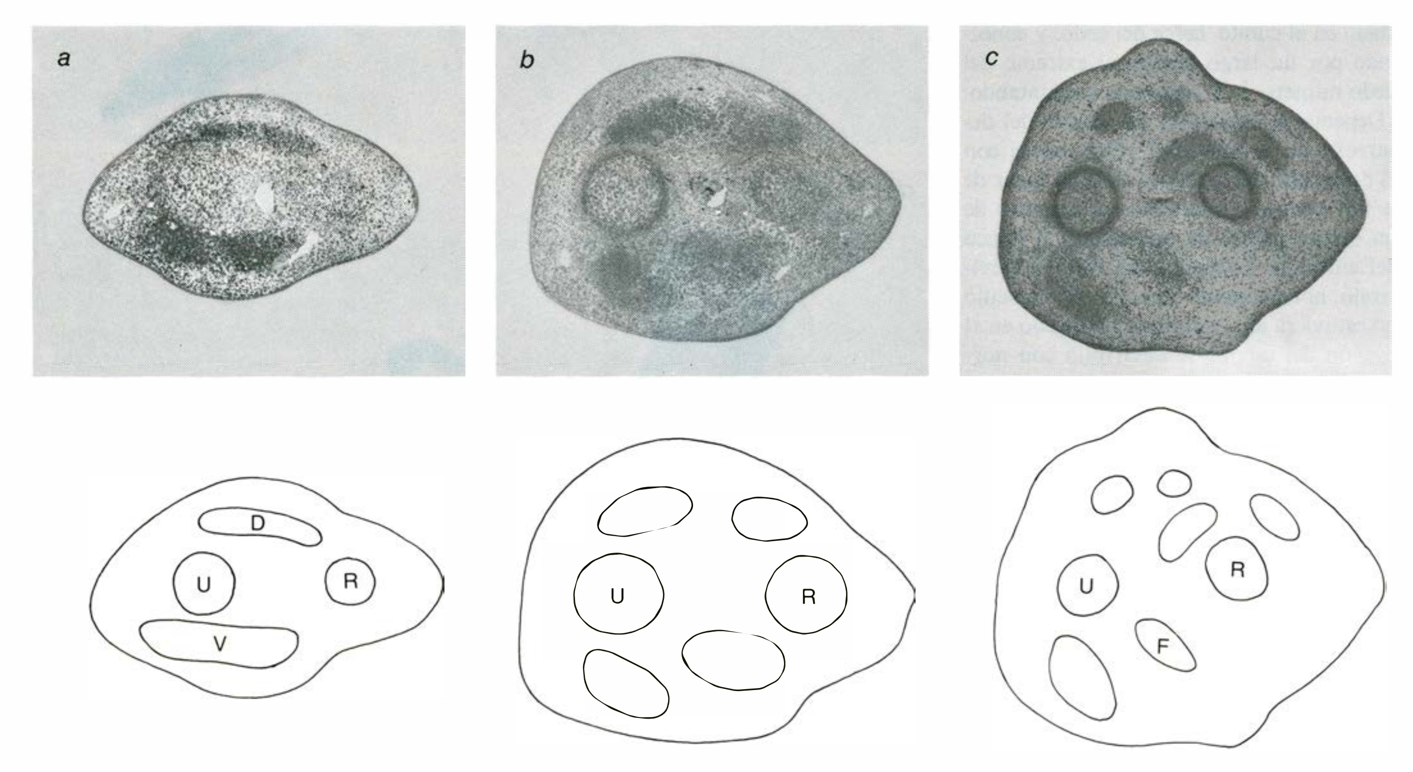


MUSCULOS ALARES de un polluelo embrionario, vistos desde abajo; comprenden uno llamado FPD (*color*), que está unido mediante un largo tendón al extremo del dedo número 3 (*izquierda*). Si, al principio del desarrollo, el extremo del esbozo se injerta en el flanco del embrión, se forma una mano más o menos normal sin la parte proximal (*derecha*). En este caso, merced al trabajo de Geoffrey Shellswell, se ha mostrado que el tendón se desarrolla autónomamente; el tendón es normal aún cuando el músculo unido a él esté ausente. La fotografía fue tomada por Summerbell, colaborador del autor del artículo.

esbozo de extremidad derecha. En este caso, los tendones ventrales del injerto se unen con los tendones huésped dorsales más próximos a ellos, y los músculos dorsales pueden integrarse en los músculos ventrales.

Creemos que el modelo de los músculos y tendones está determinado por el mismo sistema de coordenadas que los elementos cartilaginosos. Otra indicación de esto es el hecho de que un injerto de la zona de actividad polarizante del tipo que ocasiona una duplicación de los huesos a modo de imagen especular causa una duplicación, en imagen especular comparable, de los músculos y tendones.

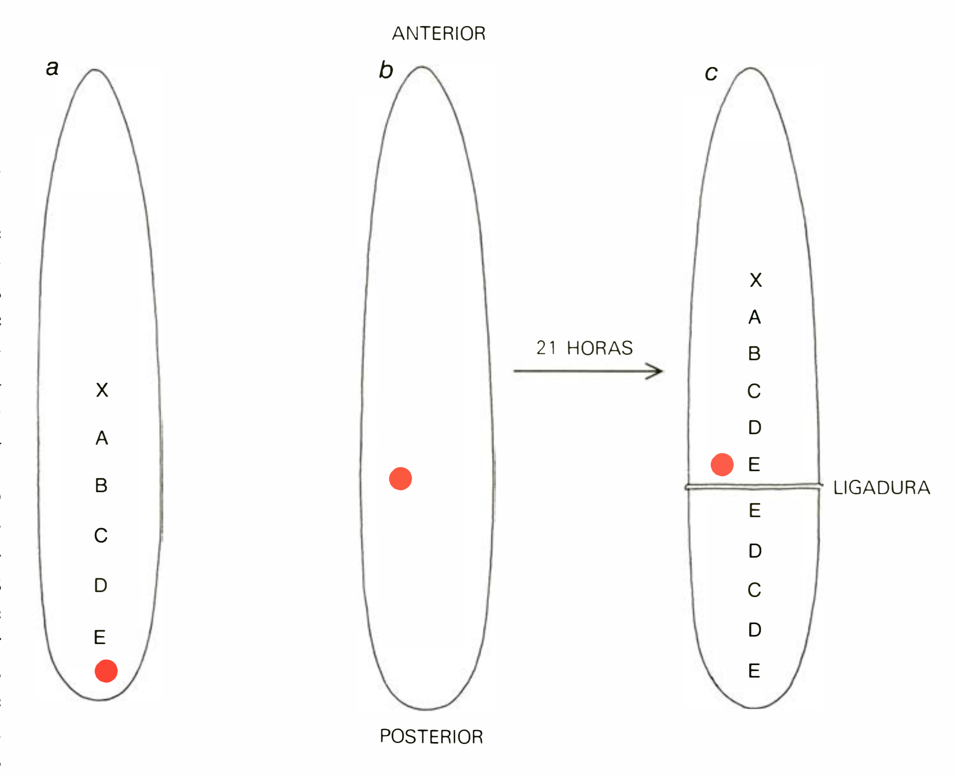
Esta discusión sobre músculos y tendones puede haber desorientado al lector al implicar que la interpretación por parte de las células de sus valores de posición es responsable no sólo del modelo de diferenciación muscular sino también de determinar qué células formarán músculo en primer lugar. De hecho la determinación de la diferenciación celular en músculo no depende de la posición en el esbozo de la extremidad. Hay una excelente prueba de B. Christ, H. Jacob y M. Jacob, de la Universidad de Bochum, y de Kieny y Alain Chevalier, de Grenoble, de que las células musculares tienen un linaje diferente del de las otras células mesenquimatosas. Migran hacia el esbozo procedentes de los somitos en un estadio precoz; aparentemente, no importa qué somito constituye su origen, lo que sugiere que determinadas células no están necesariamente destinadas a formar determinados músculos. A diferencia de las células del cartílago, las células musculares son "equivalentes". Si las células musculares están en efecto prediferenciadas, ¿qué es lo que determina el modelo de un músculo en particular? Para comprender cómo puede conseguirse esta determinación es necesario comprender ante todo cómo se va desarrollando un músculo individual. En primer lugar el músculo se hace visible en la extremidad bajo la forma de grandes bloques dorsales y ventrales, que a continuación sufren una serie de divisiones para dar lugar



MUSCULOS INDIVIDUALES, que empiezan a desarrollarse tal como se ve en las microfotografías, en las cuales se aumenta unas 35 veces una sección transversal de la extremidad. Al principio (*a*), las células musculares están agrupadas en grandes bloques musculares dorsales (*D*) y ventrales (*V*) por encima y por debajo del cúbito (*U*) y del radio (*R*). Unas 12 horas más tarde, cada bloque se ha escindido en dos (*b*). Medio día después, las masas dorsales se han dividido de nuevo; las masas ventrales lo harán más tarde. El músculo FPD surgirá de la división de una masa ventral (*F*). (Fotografías de Shellswell.)

a músculos individuales. Los primeros bloques están hechos de células musculares precursoras y de células precursoras del tejido conjuntivo. Nuestra sugerencia es que las células musculares son todas equivalentes y que el modelo de los músculos está determinado por las células del tejido conjuntivo. De acuerdo con este enfoque, las células conjuntivas de las regiones donde toman forma los bloques musculares tienen propiedades tales que las células musculares precursoras que inmigran, procedentes del somito, tienden a adherirse mejor a estas células conjuntivas que a otras; y por ello se acumulan para formar los bloques dorsales y ventrales. Las propiedades de las células del tejido conjuntivo cambian con el tiempo, presumiblemente de acuerdo con su interpretación de la información de posición. Si las células del, digamos, centro del bloque ventral llegan a tener una afinidad inferior para las células musculares, entonces las células musculares migrarán apartándose de ellas, subdividiendo en consecuencia al músculo. Es como si las futuras células del tejido conjuntivo actuaran a modo de patrón para las células musculares.

El proceso es análogo al movimiento y modelado de las células mesenquimatosas en el embrión del erizo de mar que Trygve Gustafson, de la Universidad de Estocolmo, y yo estudiamos hace algunos años. Allí las células en movimiento podían observarse directamente mediante imágenes tomadas a intervalos, y era evidente para nosotros que los pseudópodos, las expansiones temporales mediante las que se mueven las células, estaban continuamente tomando contacto con el substrato para separarse de éste después; los modelos cambiantes en la distribución de las células parecían reflejar una estabilidad cambiante en su contacto con el substrato, de modo muy similar a como el modelado del músculo en la extremidad del polluelo parece reflejar contactos entre las células musculares y un substrato de tejido conjuntivo. El modelado de los músculos por el tejido conjuntivo tendría importantes implicaciones para el modelado de los nervios en una extremidad en desarrollo y en cómo se conectan con los músculos. No sabemos cómo se establecen las conexiones correctas entre las células nerviosas y los músculos, pero sí está implicado algún tipo de especificidad; ésta puede residir más en el tejido conjuntivo que en las propias células musculares.



HUEVO DE SALTAMONTES (*a*). Tiene una región en el extremo posterior (*punto coloreado*) de efecto muy similar al de la zona de actividad polarizante en la extremidad del polluelo. El huevo del insecto origina los segmentos embrionarios del *A* al *E* (la región *X* no da lugar a nada en el embrión propiamente dicho). Desplazando la región a una posición más anterior (*b*) y ligando luego el huevo, Klaus Sander, de la Universidad de Friburgo, obtuvo embriones con un nuevo modelo de segmentos (*c*) de forma parecida al modelo de los dedos en el esbozo del ala del polluelo, cuando una zona se desplazaba hacia el centro.

Hay ahora una prueba substancial de que los gradientes en la información de posición son la razón fundamental de la formación de modelos en diversos organismos. Parte de la prueba es antigua; el papel de los gradientes al principio del desarrollo del erizo de mar, por ejemplo, fue propuesto por los investigadores suecos John Runnström y Sven O. Hörstadius en los años 1920 y 1930. Las pruebas recientes resultan particularmente impresionantes en el caso de los insectos. Tanto Peter A. Lawrence, del British Medical Research Council's Laboratory of Molecular Biology en Cambridge, como Horst Bohn, del Instituto Zoológico en Munich, han informado que el modelo de la epidermis de los insectos puede entenderse en términos de gradientes. Es de particular importancia el reciente descubrimiento de compartimientos en el desarrollo de los insectos por Antonio García-Bellido, de la Universidad de Madrid. Son grupos de células en una posición especial cuyo programa de desarrollo cambia al mismo tiempo. Una sucesión de tales acontecimientos divide el ala embrionaria, por ejemplo, en un cierto número de regiones distintas. Los genes asociados con la formación de compartimientos pueden identificarse y manipularse, de modo que el estudio de este proceso constituye un prometedor enfoque de las relaciones entre la acción de los genes y la formación de modelos.

Klaus Sander, de la Universidad de Friburgo, ha demostrado cómo los gradientes pueden controlar el modelaje más precoz en un huevo de insecto en desarrollo. Parece que el margen posterior del huevo de saltamontes puede considerarse fuente de un gradiente muy similar a la zona de actividad polarizante que se halla en la extremidad del polluelo. Las similitudes formales de los dos sistemas son sorprendentes. El desplazamiento de parte del citoplasma del extremo posterior del huevo hacia el interior puede dar lugar igual que el modelo de dedos obtenidos injertando una zona adicional de actividad polarizante en esbozo joven de ala. Tanto el citoplasma posterior del huevo como la zona de actividad polarizante del esbozo del ala parecen actuar como regiones frontera que proporcionan una señal de posición. Regiones análogas se encuentran en otros animales: el clásico "organizador" en los embriones de los anfibios, la cabeza de la hidra y los micrómeros en los embriones de los erizos de mar.

Los mecanismos basados en la información de posición no son los únicos por los cuales puede determinarse el modelo durante el desarrollo. Hay otro mecanismo que implica la presencia de determinantes de algún tipo en lugares específicos dentro del citoplasma del huevo: al dividirse el huevo, las células de diferentes regiones adquieren diversos componentes citoplamáticos que especifican el curso de su diferenciación. Karl Illmensee y Anthony P. Mahowald, de la Universidad de Indiana, encontraron que lo que se llama plasma polar, en el extremo posterior del huevo de insecto, determina el desarrollo de células germinales en el insecto. Al

transplantar el citoplasma al extremo anterior del huevo provocaron que las células que normalmente hubieran dado lugar a estructuras cefálicas se convirtieran en células germinales. La localización citoplasmática de este tipo constituye la excepción más que la regla, por lo menos para los vertebrados, y no explica la mayoría de modelos existentes. Un gran número de experimentos indica que la mayoría de modelos surgen como resultado de la interacción intercelular; si se extraen o cambian de sitio ciertas partes del embrión joven, alguna forma de regulación arregla las cosas de modo que se desarrolle un animal normal.

Dicha regulación del modelo cuando se extrae una parte durante el desarrollo, así como la regeneración de una estructura perdida en un animal adulto, pueden entenderse en términos de información de posición. La regulación y la regeneración reclaman el reemplazo de valores de posición que faltan por otros nuevos valores, que son interpretados de nuevo por las células. Hay dos maneras principales en que esto puede ocurrir, según observó ya Thomas Hunt Morgan a principios de siglo. En la morfalasis (como en la regeneración de la hidra o en la mayoría de las regulaciones dentro del embrión joven) la estructura regenerada se moldea a partir del tejido restante sin la proliferación de células. La morfalasis puede comprenderse como el establecimiento de una nueva región frontera en la superficie cortada y la determinación de nuevos valores de posición con respecto a esta frontera. En la epimorfosis (la regeneración de la extremidad de los anfibios constituye un ejemplo clásico) la estructura regenerada surge mediante el crecimiento: la proliferación de un pequeño grupo de células en la superficie cortada. En el nuevo tejido se generan nuevos valores de posición.

Los trabajos de Peter J. Bryant y Susan V. Bryant, de la Universidad de California en Irvine, y Vernon French, de la Univer sidad de Edimburgo, han arrojado nueva luz. Desarrollaron un modelo basado en un sistema coordenado polar, en el cual el valor de posición de cualquier célula viene definido por dos coordenadas, una meridional y otra a lo largo del eje proximodistal. Una característica importante de su modelo es que, intercalada o interpoladamente, la regeneración ocurre siempre que valores de posición discordantes se hacen adyacentes uno al otro; se generan nuevos valores de posición en el tejido en crecimiento hasta que se elimina la discordancia. Estos investigadores han mostrado que su modelo puede servir para multiples fenómenos (en las extremidades de los insectos y anfibios y en los discos imaginables de las larvas de los insectos). Resulta

alentador que los mismos tipos de fenómenos sean observados en diversos animales y que obedezcan a las mismas reglas, si se cree en la universalidad de los mecanismos para la formación de modelos.

He intentado mostrar que el concepto de información de posición proporciona un útil marco de trabajo dentro del cual puede comprenderse e investigarse el desarrollo y la regulación de los modelos biológicos. Hay abundancia abrumadora de pruebas (algunas de las más impresionantes corresponden a estudios sobre el desarrollo de los insectos, que aquí sólo hemos citado) sobre el hecho de que las células presentan una propiedad que corresponde a su valor de posición y que las reglas para la formación de modelos y regulación son muy similares en sistemas muy diferentes. Si tales afirmaciones son correctas, los que trabajamos en estos problemas estamos en la misma situación en que se encontraban los genetistas antes de que se identificara el ADN como el material genético: tenemos reglas que gobiernan la fenomenología pero la base molecular de los fenómenos nos resulta completamente desconocida. No tenemos idea de cómo se lleva a cabo la señalización de posición (podrían estar implicados simples iones o complejos ácidos nucleicos). Además, si la información de posición es similar en sistemas diferentes, la amplia gama de modelos debe surgir de la diferente forma de interpretar los valores de posición por parte de las células. Un problema central es, obviamente, saber cómo debe interpretarse la información.

A veces se sostiene que no se ha hecho ningún progreso real mientras no se afinque un mecanismo biológico sobre una base molecular firme (en este caso, hasta que no se conozca la naturaleza molecular de los gradientes, o señales de posición). Dicha visión niega la existencia de niveles diferentes de organización en los que pueden investigarse los procesos biológicos con conocimiento de causa. A los biólogos del desarrollo les gustaría conocer la base molecular de la formación de modelos, pero de momento no hay una forma obvia de encontrarla. Nuestros problemas son similares a los de cualquier investigador que intenta reducir el comportamiento celular a términos moleculares, sobre todo cuando implica reacciones dentro de las células que dan lugar a cambios en el estado o en el comportamiento celular. Quizá deberíamos ser menos beligerantes y recordar que el estudio de la genética era (y es) eficaz a otros niveles además del nivel de ADN, y que, mientras no se disponga de la correcta fenomenología, no sabemos lo que estamos intentando explicar o dónde buscar la explicación.

Una migración en masa de aves terrestres sobre el océano

Cada otoño, millones de pájaros y pequeñas aves litorales abandonan la costa oriental de Norteamérica y se desplazan hacia el Caribe y Sudamérica. Su difícil viaje se ha seguido con la ayuda del radar

Timothy C. Williams y Janet M. Williams

Durante los dos meses que siguen a la partida de los veraneantes de las playas de Nueva Inglaterra y de las Provincias Marítimas del Canadá, estas regiones experimentan una marcha mucho más nutrida que pasa inadvertida a la mayoría de los habitantes que se quedan. Más de 100 millones de aves, la mayoría muy pequeñas, cruzan estas costas en su vuelo anual hacia el sur, sobre el Atlántico y desde Norteamérica oriental hacia el Caribe y Sudamérica. Un observador paciente, provisto de una par de potentes prismáticos, puede ver algunas de estas aves cuando pasan por delante del disco de la luna llena o atraviesan un haz de luz dirigido hacia arriba, durante la noche; pero en su mayor parte los participantes en la migración no pueden verse directamente. Con la ayuda de John M. Teal y John W. Kanwisher, de la Woods Hole Oceanographic Institution, de Leonard C. Ireland, de la Bermuda Biological Station, y de otros investigadores hemos utilizado una red de radares para observar el comportamiento de las aves. Esos estudios han proporcionado mucha información acerca de esta notable migración y han indicado porqué algunas aves tienen éxito en ella y otras, que parecen hallarse sobre el océano por error o por infortunio, perecen en gran número.

Las aves que hemos observado mediante el radar a lo largo de la costa atlántica durante la última semana de septiembre y la primera mitad de octubre son, sobre todo, aves litorales (como correlimos, andarríos, chorlitos y chorlitejos) y pequeños paseriformes (como currucas americanas: picafollas, dendroicas, etcétera). La inmensa mayoría de aves empiezan su viaje de noche, después de que un sistema atmosférico de frente frío se haya desplazado en dirección sudeste sobre la costa. Quien se hallára en una zona tranquila del cabo Cod en las primeras horas de una noche de octubre y con un fresco viento del noroeste a su espalda, podría oír a numerosas avecillas llamándose entre sí mientras pasan sobre su cabeza. Si este observador se dirigiera entonces a cualquiera de las tres grandes instalaciones de radar del cabo (vigilancia de la Fuerza Aérea, control del tráfico aéreo o radar meteorológico) encontraría la porción central de la pantalla de radar ocupada por pequeños puntos brillantes que van desplazándose en una dirección aproximadamente sudeste.

Esta pantalla de radar representa un mapa de la zona prospectada por el aparato; el norte está arriba, y la distancia de un objeto a las instalaciones de radar está indicada mediante anillos concéntricos de enfoque situados a intervalos que representan unos 10 kilómetros. La localización de cualquier objeto detectado se representa sobre la pantalla mediante un punto brillante. Los radares a los que tuvimos acceso exhibían así cada objeto de una a 15 veces por minuto.

Con el fin de registrar información a partir de la pantalla de radar tomamos fotografías con exposición diferida. Cualquier objeto que se mueva lentamente, como un ave, queda registrado en la fotografía como una serie de pequeños puntos o como un trazo. Cerrando brevemente y después volviendo a abrir el obturador al final de la exposición, podíamos determinar la dirección de movimiento del ave. Midiendo la longitud del trazo, y por tanto la distancia recorrida, podíamos calcular la velocidad del ave (conociendo la duración de la exposición). El número de aves detectadas por unidad de superficie sirvió como base para estimar la densidad de migración, que clasificamos en cuatro categorías: sin movimiento, migración débil, migración moderada y migración intensa. La mayoría de aparatos de radar podían estimar asimismo la altitud de las aves a partir del ángulo de elevación del haz radárico y de la distancia de las aves al radar.

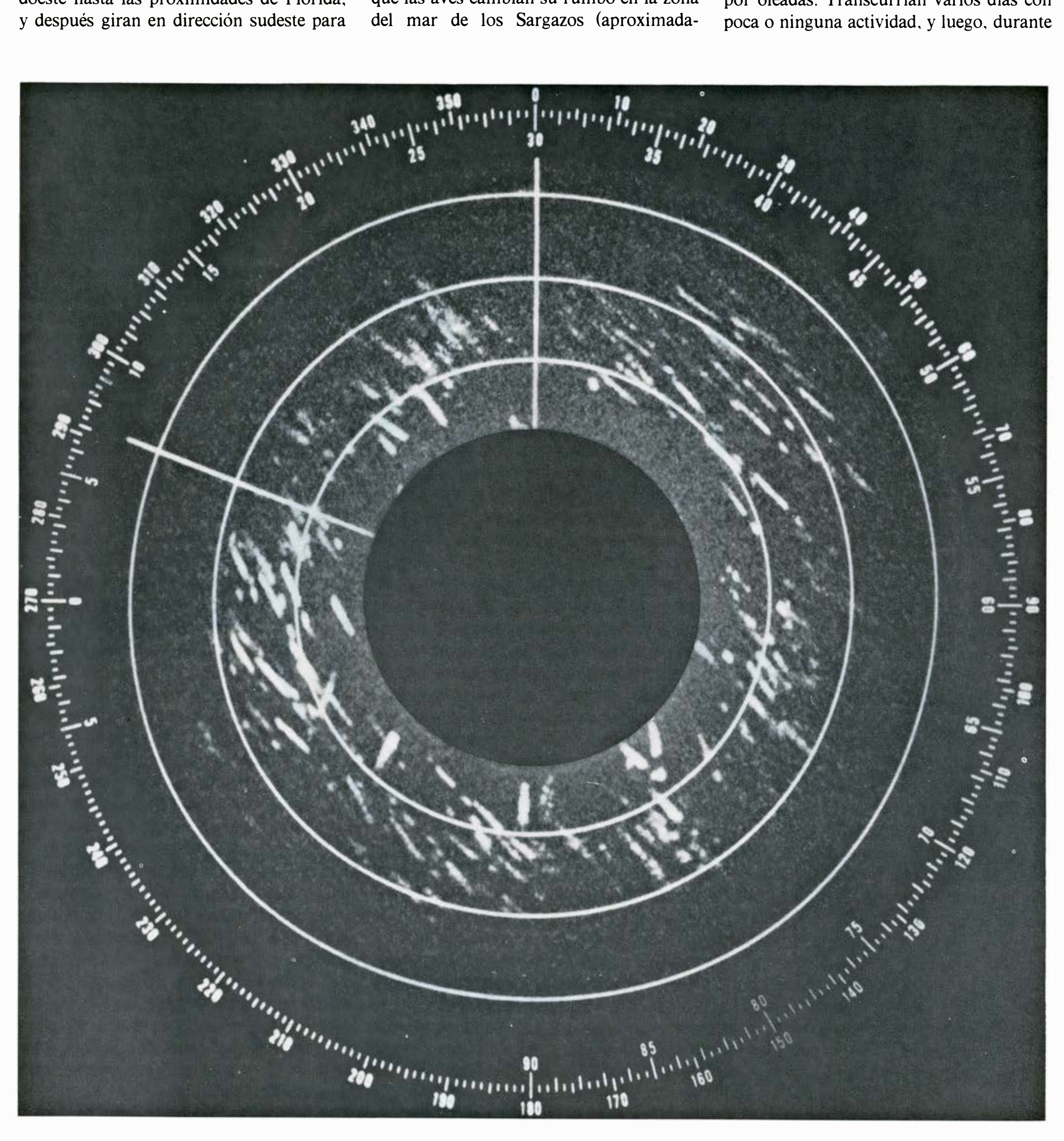
Empezamos nuestra investigación sobre estas migraciones en dirección al sur al observar las aves mientras éstas se desplazaban sobre el cabo Cod. Como les ocurriera a William H. Drury y a Ian C. T. Nisbet, de la Massachusetts Audubon Society, antes que a nosotros, quedamos fascinados por el gran número de aves (hasta 12 millones por noche) que abandonaban el cabo Cod dirigiéndose hacia el sudeste, en pleno Atlántico. Para determinar qué les pasaba a estas aves, las observamos desde radares situados en las islas de Bermuda y Antigua. Descubrimos que muchas aves, pero no todas, continuaban volando hacia el sudeste sobre Bermuda, pero llegaban al Caribe volando en dirección sudoeste. Las aves terrestres y la mayoría de aves litorales no habían tenido ninguna posibilidad de descender a tierra durante este vuelo de más de 3000 kilómetros. Además, no habían podido navegar siguiendo las líneas de costa ni nigún otro punto de referencia; era preciso algún otro sistema de dirección.

El único modo en que podíamos estudiar el movimiento de las aves mientras éstas se hallaban sobre el agua era mediante radares en lugares costeros y en islas y barcos situados a lo largo de la ruta. Estas observaciones tenían que hacerse simultáneamente, con radares similares y bajo condiciones parecidas. Por tanto, durante la última semana de septiembre y las dos primeras semanas de octubre, por seis años, nos procuramos la cooperación de seis gobiernos nacionales, cuatro oficinas federales, la Woods Hole Oceanographic Institution y gran número

de amigos nuestros y estudiantes para manejar hasta nueve estaciones de radar a lo largo de la ruta de migración.

Los modelos básicos de migración que encontramos parecen comprender dos rutas, por lo menos, desde Norteamérica a Sudamérica. En el curso de una de ellas, las aves siguen la costa en dirección sudoeste hasta las proximidades de Florida, y después giran en dirección sudeste para desplazarse a lo largo de las islas del Caribe. La segunda ruta es la que toman las aves que abandonan la costa desde Nueva Escocia, al norte, hasta Virginia, al sur, y después se desplazan en dirección sudeste (dirección de vuelo que se observa desde barcos situados en todas las zonas del Atlántico que estudiamos). Suponemos que las aves cambian su rumbo en la zona del mar de los Sargazos (aproximadamente el límite sur de nuestras observaciones), al encontrarse con los vientos alisios del nordeste. Ayudados por estos vientos, se desplazan en dirección sudoeste sobre la región que ocupan las islas del Caribe.

En todos los emplazamientos de radar descubrimos que la migración se realiza por oleadas. Transcurrían varios días con poca o ninguna actividad, y luego, durante



LAS AVES MIGRADORAS aparecen como trazos blancos en esta fotografía de una pantalla de radar en la isla caribe de Antigua. El norte está arriba. Los círculos blancos representan intervalos de 10 kilómetros desde la estación de radar. El centro de la pantalla está oscurecido para evitar la sobreexposición de la fotografía debida a las señales procedentes del suelo. La fotografía está tomada con una exposición de cinco minutos, y en ella la mayor concentración de trazos que representan aves se encuentra en los círculos situados entre los 20 y los 45 kilómetros. El punto que se observa al final de un trazo, originado al cerrar brevemente el obturador de la cámara, indica la dirección del movimiento. La línea blanca de la izquierda muestra el ángulo de elevación del haz de radar, que interceptaba una capa de aves situada a una altitud de tres a seis km. Se observó que la mayoría de las aves se desplazaban en dirección sudeste.



MIGRADORES ATLANTICOS, dibujados a la misma escala. Se incluyen tres aves litorales: la aguja del Hudson (*a*), el chorlito dorado chico (*b*), el correlimos de Bonaparte (*c*) y un pájaro, el picafollas estriado (*d*). Quizá la mitad de migradores son paseriformes. La curruca da una indicación de la escala: mide de 11,5 a 14 cm y pesa menos de 30 gr. Empieza su viaje de noche, después de que en un frente frío se halla desplazado en dirección sudeste.

uno o dos días, las aves se desplazaban en gran número. Los intervalos entre los períodos de actividad migratoria se hacían más largos conforme nos alejábamos de la costa. En las estaciones costeras casi siempre se podía registrar algún movimiento durante la noche. En Bermuda encontramos períodos ocasionales sin migración, y en el Caribe solían pasar varios días sin que se observara ningún migrante.

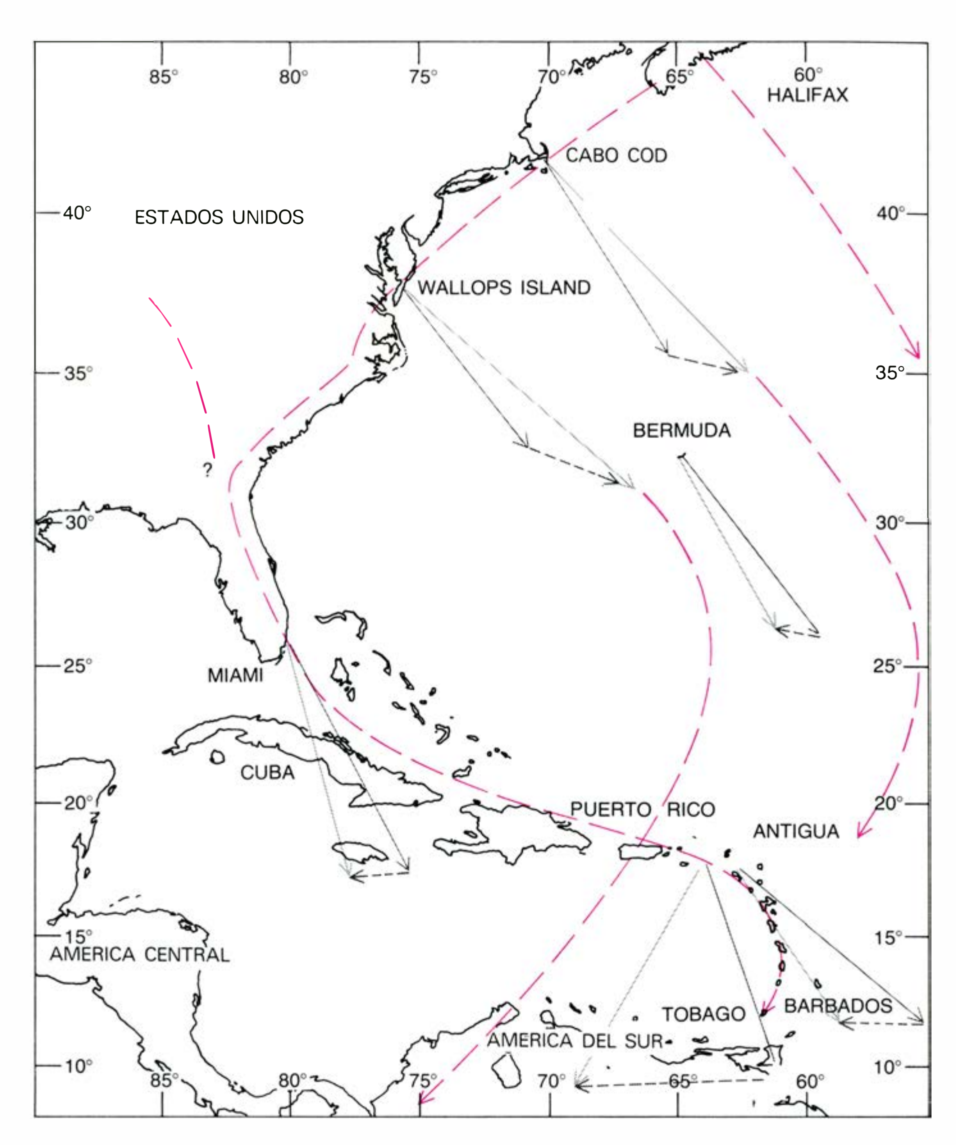
Al superponer los datos obtenidos mediante el radar sobre una serie de fotografías hechas desde satélites meteorológicos operados por la National Oceanic and Atmospheric Administration (NOAA) puede establecerse una correlación entre la actividad migratoria y el estado del tiempo. Un ejemplo lo proporcionan las fotografías desde satélite correspondientes a los días 3 a 6 de octubre de 1973 [*véase la ilustración de la página siguiente*]. El primer día empezó una partida de aves desde la costa norteamericana. Un frente frío acababa de moverse mar adentro desde el cabo Cod a Florida. Como muestra la disposición de las nubes, el frente cruzó la línea de costa entre el cabo Cod y Halifax. Los radares costeros registraron intensos movimientos de aves esa noche; las aves se desplazaban a lo largo de la costa y mar adentro.

Hacia el cuatro, de octubre, el sistema frontal se hizo estacionario entre Bermuda y la costa. Las aves observadas desde un barco penetraron durante todo el día en el sistema frontal, desplazándose hacia el sudeste y alcanzando la zona de Bermuda a media tarde. Al día siguiente, 5 de octubre, estas aves se encontraban ya entre Bermuda y el Caribe. En Bermuda, el movimiento había dejado de tener dirección sudeste y era hacia el sur (lo que posiblemente reflejaba el movimiento general de las aves que se habían desplazado desde Halifax la noche anterior). Los movimientos de las aves vistas desde el barco, que se hallaba a unos 400 kilómetros detrás (al noroeste) del frente atmosférico, parecían dispersos en todas direcciones, aunque la densidad de las aves se clasificó como elevada.

El 6 de octubre la oleada de la migración alcanzó el Caribe, con movimientos intensos registrados en Antigua durante el día y en Barbados durante el atardecer. El tiempo total de vuelo desde Norteamérica a Antigua fue de unas 64 horas y, a Barbados, de unas 72 horas.

El calendario de esta migración es el típico de los datos que hemos podido reunir. El disparo de salida de la costa lo dan las condiciones de vuelo favorables después de un frente frío. El carácter significativo del tiempo lo constituyen los fuertes vientos del noroeste, que ayudan a las aves a desplazarse desde la costa hasta la zona de Bermuda en un tiempo medio de sólo 18 horas. El radar indica que son pocas las aves que aterrizan en Bermuda; la gran mayoría continúa en dirección sudeste sobre la isla, por lo general con vientos suaves y variables, hasta que alcanzan la zona del mar de los Sargazos, donde encuentran los vientos alisios del nordeste que les ayudan a llegar el Caribe. Esta segunda parte del viaje es mucho más lenta que la primera: las aves parecen estar volando durante unas 48 horas entre Bermuda y Antigua. Los radares de Antigua indican que, como en Bermuda, pocas aves toman tierra; en lugar de ello la mayoría parecen continuar volando durante otras 18 horas hasta alcanzar Sudamérica. El tiempo total de vuelo sin escalas sobre el océano, según se desprende, es pues de unas 86 horas, con un margen de variación de 12 horas en más o en menos.

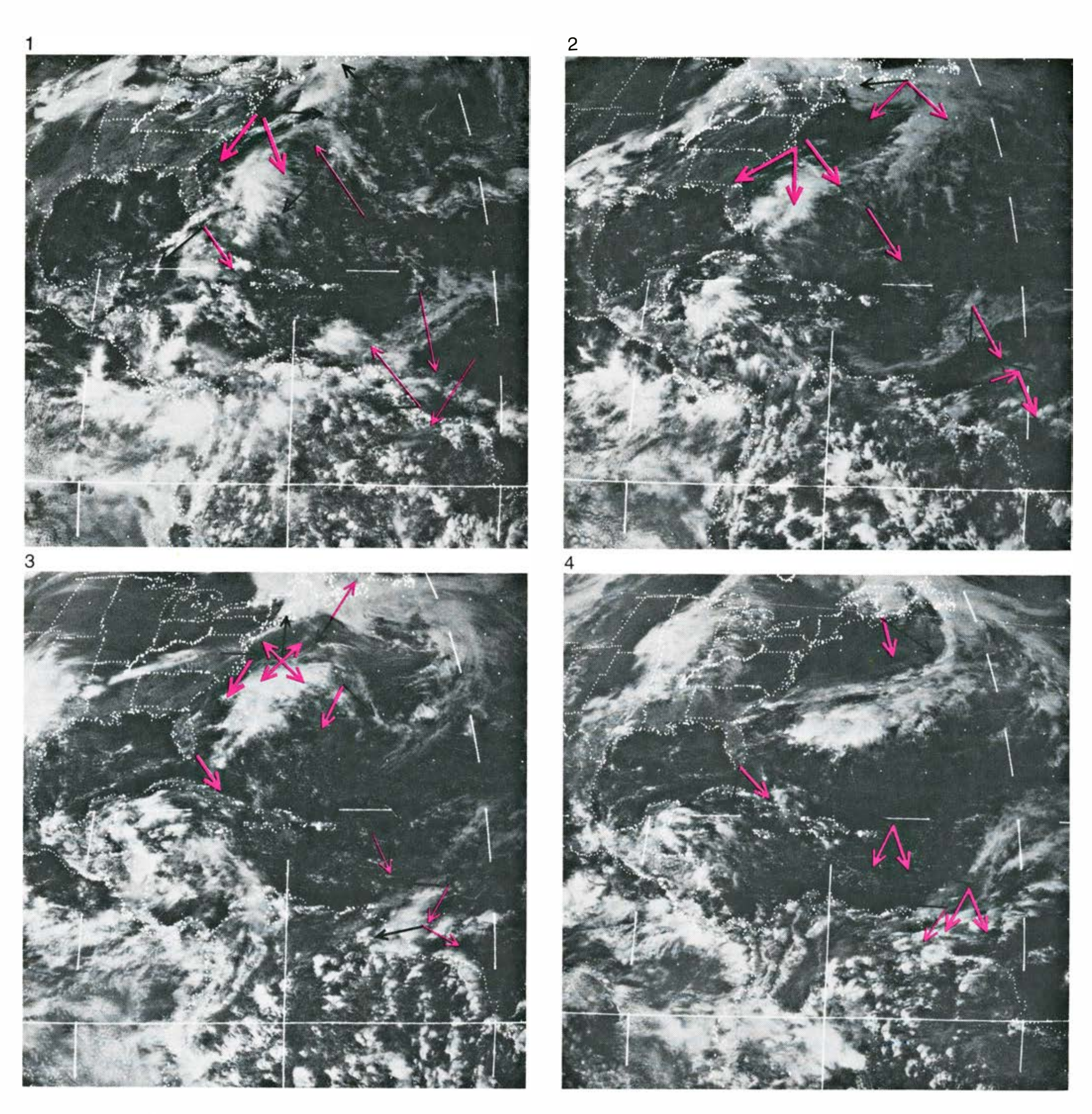
Hasta donde hemos podido averiguar, esta migración constituye el vuelo sin escalas más largo (tanto por la duración como por la distancia) que se conoce para aves pequeñas. Asimismo, parece tener lugar a una altitud sobre el suelo superior a la de ninguna otra migración aviar. A partir de las observaciones del radar descubrimos que las migraciones más densas



ZONA DE MIGRACION desde Norteamérica al Caribe y a Sudamérica. El mapa está basado en observaciones de radar desde Halifax, cabo Cod, Wallops Island, Bermuda, Miami, Puerto Rico, Antigua, Barbados y Tobago. Las líneas de color discontinuas indican dos series de posibles rutas migratorias, una para las aves que vuelan a lo largo, o cerca, de la costa de América del Norte y la otra para las aves que realizan la mayor parte del viaje sobre el océano. (Los datos para Puerto Rico proceden de observaciones realizadas por W. J. Richardson, de Environmental Research Associates, de Toronto.) Las líneas grises que forman triángulos señalan la relación del viento al rumbo y a la ruta de las aves. En cada caso la línea de trazos corresponde a la dirección del viento (la velocidad relativa del viento viene indicada por la longitud de la línea), la línea gris obscuro representa el rumbo medio de las aves y la línea gris claro muestra su curso medio. Las aves mantienen de forma constante un rumbo sudeste, pero los vientos alisios que soplan del nordeste en el Caribe crean una deriva que tiene por efecto desviar a las migradoras al sudoeste cuando se acercan a su destino.

desde la costa de América del Norte a Bermuda se encuentran a una altitud de dos kilómetros o algo inferior, aunque se detectan algunas aves a cinco kilómetros. En Bermuda da comienzo un modelo distinto. La mayoría de las aves vuela a una altitud de uno a dos kilómetros. Cuando llegan a Antigua se hallan entre tres y seis kilómetros, y algunos días el radar mostraba un número significativo de aves a 6,5 kilómetros de altitud. Aunque hicimos muchas menos observaciones en Barbados y Tobago, parece que entonces las aves estaban descendiendo, hasta que en Tobago no observamos ningún ave por encima de los 800 metros. Evidentemente, al acercarse a la costa de América del Sur las aves van bajando gradualmente mientras se preparan para tomar tierra.

Las aves situadas a una altitud de seis kilómetros sobre Antigua están volando a una temperatura del aire de cero grados Celsius, y el aire sólo contiene allí aproximadamente la mitad del oxígeno que po-

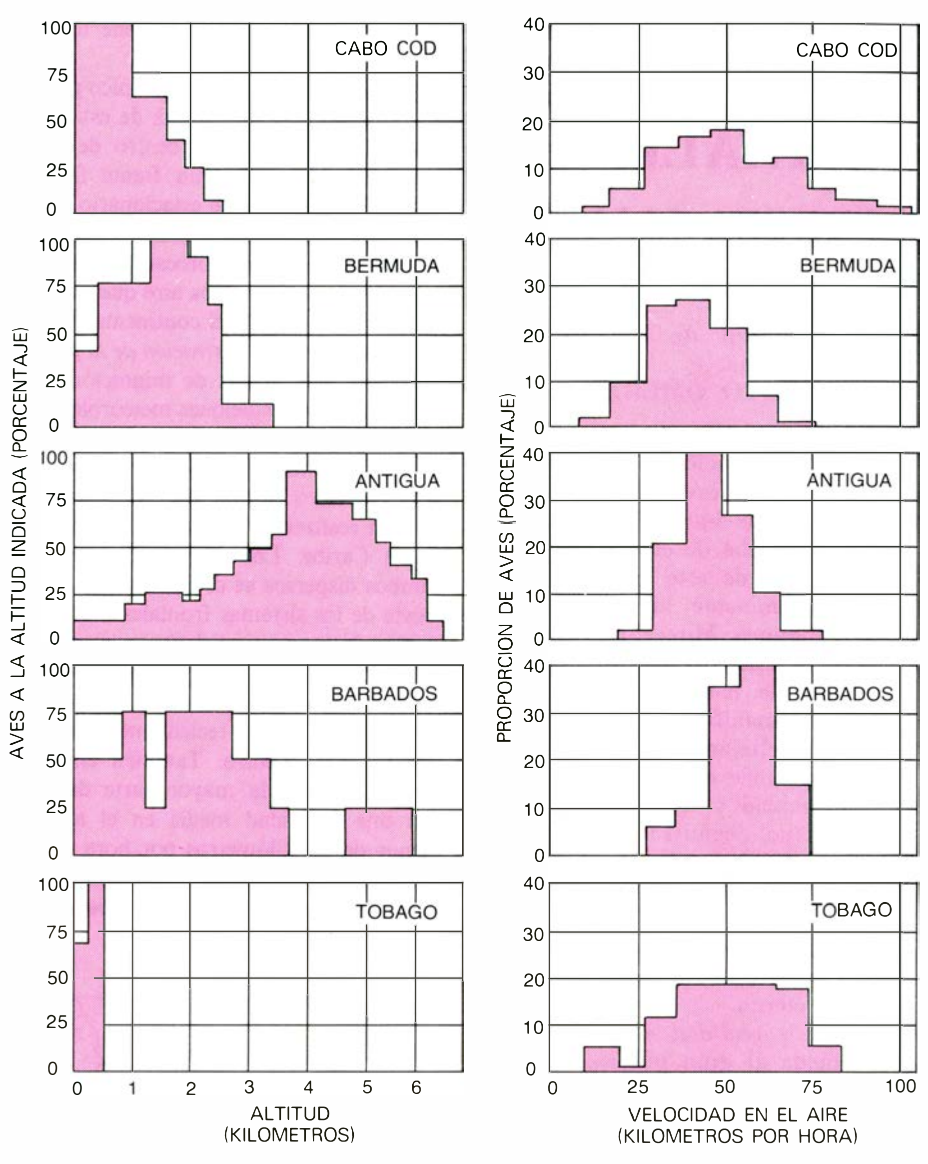


MARCHA DE LA MIGRACION durante cuatro días de octubre de 1973. Tal avance se ha cartografiado sobre unas fotografías de satélite meteorológico realizadas por la National Oceanic and Atmospheric Administration (NOAA). Las flechas de color representan el curso medio de una migración aviar, y el grosor de la flecha indica si la migración era débil, media o intensa. Las flechas negras señalan la dirección del viento a la altitud de las aves; cuanto más larga es la flecha, mayor es la velocidad del viento. Una migración empezó el 3 de octubre en condiciones típicas (*1*), después de que un frente frío acababa de desplazarse mar adentro desde el cabo Cod hasta Florida. Los vientos acompañantes del noroeste ayudaron a las aves. Al día siguiente (*2*), el sistema frontal se hallaba estacionario entre Bermuda y la costa; la mayoría de las aves habían alcanzado Bermuda. Después de sobrepasar Bermuda (*3 y 4*), encontraron los vientos alisios del nordeste que las hicieron cambiar de dirección, dirigiéndolas hacia su destino en el Caribe y América del Sur. Las aves se encontraban en la isla de Barbados (en las Antillas menores) al cuarto día, después de unas 72 horas de vuelo. El tiempo de vuelo sin escalas sobre el océano duró más de 86 horas.

see al nivel del mar. La ventaja de volar a estas alturas parece consistir en que las aves se sitúan en una región con vientos favorables. Calculamos la velocidad media del viento a distintas altitudes en los días en que detectamos migraciones moderadas o intensas en Antigua a una altitud media de 4,2 kilómetros. La dirección media del movimiento de las aves en estos días era sudeste; nuestros datos para altitudes inferiores a cuatro kilómetros mostraban fuertes vientos del este-sudeste, que las aves que volaban a gran altura evitaban.

En el aspecto fisiológico, es posible el vuelo a estas altitudes merced al especial sistema respiratorio de las aves. Este sistema encierra distintas adaptaciones para el vuelo, pero en nuestro caso la más importante es el flujo de aire y sangre a contracorriente en el pulmón del ave. Representa la clave de la capacidad del animal para obtener oxígeno de manera eficiente, y, por tanto, para poder volar a altitudes elevadas.

Observaciones hechas tanto visualmente (mediante prismáticos) como a través del radar desde barcos en alta mar proporcionaron una idea mucho más clara del modo de actuar de estas aves durante su largo vuelo sobre el océano, e indicaron asimismo qué tipos de aves podían estar realizando el vuelo. Para seguir nuestro análisis de tales resultados debe quedar clara la distinción entre la ruta y el rumbo de un ave. Si un ave está volando en un viento, el movimiento que se aprecia en la pantalla de radar (la ruta o curso) no corresponde necesariamente a la dirección en la que el ave orienta su cuerpo (el rumbo). El viento introduce una deriva, es decir, hace que el ave se desplace hacia el lado de su rumbo situado a favor del viento.



ALTITUD Y VELOCIDAD de las aves migradoras, según se desprende de las observaciones por radar. El gráfico de altitud (*izquierda*) muestra el porcentaje de aves que se encontraba en las altitudes indicadas durante los momentos de migración moderada o intensa. La altitud media alcanza un máximo en Antigua y después disminuye al acercarse las aves a su destino. El gráfico de velocidad en el aire (*derecha*) refleja el porcentaje de aves detectado a cada una de las velocidades. La pérdida de las aves más rápidas y de las más lentas entre el cabo Cod y Antigua sugiere que sólo ciertas especies están adaptadas a tan larga migración. El número de ellas que vuela a menos de 50 kilómetros por hora indica que muchas son pequeños paseriformes.

Otro factor que debe tenerse en cuenta es la velocidad del ave en el aire, que es naturalmente la velocidad de su vuelo en el aire. La velocidad en el aire y el rumbo se calculan, a partir de los datos de radar, mediante la dirección y la velocidad del viento, determinadas por un globo atmosférico situado a la altitud del ave. Con el fin de distinguir el comportamiento de las aves con respecto a la deriva causada por los vientos, cartografiamos la ruta y el rumbo para las observaciones de migraciones de aves en el mar. Siguiendo el ejemplo de Ronald P. Larkin, de la Universidad de Rockefeller, que analizó los dos primeros años de estas observaciones en el mar, dividimos todas las observaciones obtenidas desde barcos en períodos de comportamiento migratorio uniforme.

Al superponer estos datos sobre un mapa aparecen cuatro categorías de comportamiento bastante distintas [*véase la ilustración de la página siguiente*]. La primera categoría consta de los períodos en los que no se detecta ninguna migración significativa. La segunda corresponde a aves cuyo rumbo es sur a sudeste y cuya ruta es hacia el sur. Estas aves, si continuaran con esta disposición, llegarían probablemente a Sudamérica. La tercera categoría está formada por grupos de aves que parecen desorientadas. Sus rumbos son muy dispares y cualquier avance neto del grupo ha de atribuirse en gran parte a la deriva causada por el viento. La velocidad en el aire calculada para estas aves es por lo general bastante baja; no llega a los 20 kilómetros por hora. Parece poco probable que una proporción significativa de las aves de esta categoría alcancen el Caribe o Sudamérica si continúan en su vuelo pobremente orientado y a velocidades bajas. La última categoría está formada por aves que aparentemente intentan desplazarse en una dirección distinta de la que les llevaría hasta el Caribe.

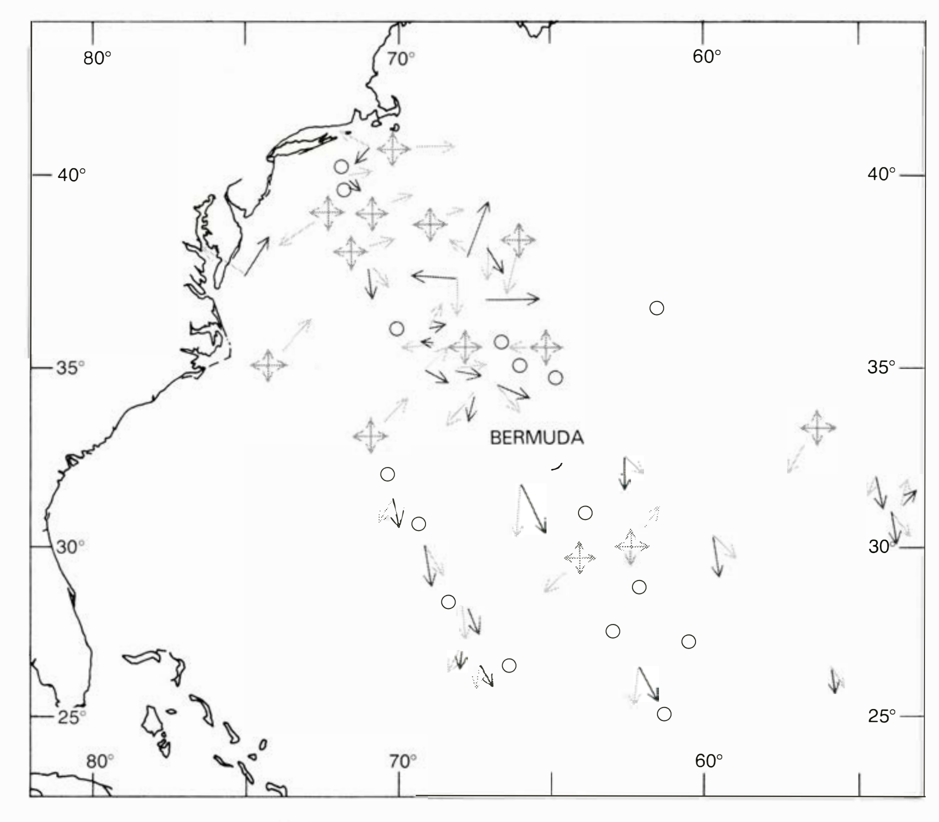
No pudimos advertir ninguna tendencia estadísticamente significativa para agrupar a estos cuatro tipos de comportamiento en ninguna de las zonas del Atlántico que estudiamos. Por ello llegamos a la conclusión de que algún otro factor que no fuera la geografía debía estar determinando las categorías de comportamiento que observamos desde los barcos. Si se cartografían los mismos datos sobre un esquema de las estructuras meteorológicas predominantes

en un momento dado se obtiene una visión más clara.

Un sistema meteorológico típico para el que hicimos una cartografía de esta clase estaba formado por un centro de bajas presiones o borrasca, un frente frío en movimiento, un frente estacionario y dos masas de aire de altas presiones, una subtropical y la otra, al noroeste del sistema frontal, que representaba aire que acababa de desplazarse desde el continente norteamericano [*véase la ilustración de la página siguiente*]. Las pautas de migración diferían según las condiciones meteorológicas. En la zona que rodeaba el anticiclón subtropical encontramos únicamente aves migradoras cuyo rumbo era sur o sudeste; estaban realizando un avance efectivo hacia el Caribe. Los grupos de aves con rumbos dispersos se divisaron sólo al noroeste de los sistemas frontales o al sur y al este de los centros de bajas presiones. En ambos casos, constituía un factor común el que las aves desorientadas acababan de experimentar recientemente fuertes vientos mar adentro. También en estas zonas hallamos la mayor parte de aves con una velocidad media en el aire de menos de 20 kilómetros por hora.

A partir de este análisis, podemos clasificar provisionalmente las aves observadas sobre el Atlántico en verdaderos migradores, si han penetrado en el sistema frontal que inició su partida, y en migradores desafortunados, cuando se observan en un centro de altas presiones al noroeste de un sistema frontal o en un centro de bajas presiones. (En la segunda categoría pueden existir ocasionalmente migradores verdaderos, por haberse hecho la observación, cerca de la costa norteamericana, en un momento en el que las aves no habían penetrado todavía en el sistema frontal). Cabe que los migradores desafortunados sean aves que normalmente migran sobre tierra pero que han sido arrastradas al mar por el viento.

Un rumbo que suele mantenerse hacia el sudeste es típico de los verdaderos migradores, independientemente de si el radar que los detecta es costero o se halla instalado en una isla o en un barco. Pero no siempre resulta fácil distinguir a los verdaderos migradores debido a la frecuente presencia simultánea de dos o más pautas de comportamiento migrador, en particular en los grandes radares costeros o insulares. Y así, en cada estación separamos grupos de aves en base a la dirección de su ruta y después calculamos el rumbo medio de cada grupo. Aunque el rumbo



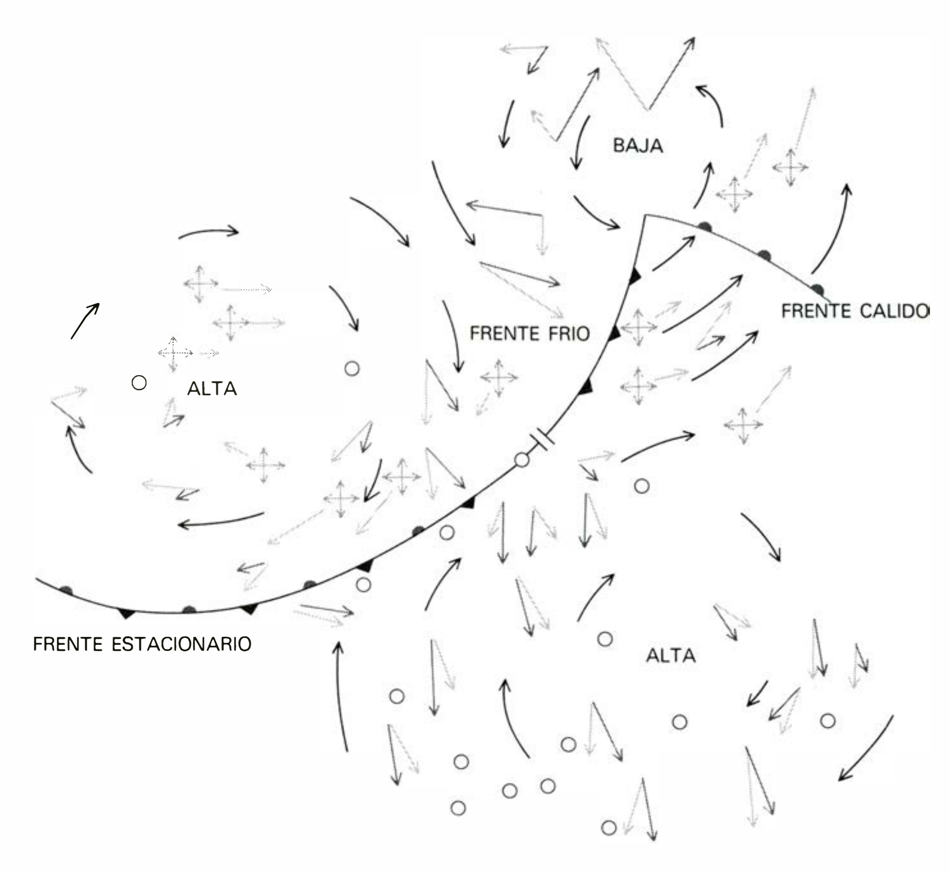
EL COMPORTAMIENTO MIGRADOR de las aves sobre el océano se observó mediante radares situados en barcos. Las flechas en gris obscuro indican el rumbo medio de las aves y, por la longitud del trazo, la velocidad relativa en el aire; cuanto más larga es la flecha, mayor es la velocidad de vuelo. Las flechas en gris claro señalan, de forma similar, el curso medio de las aves y su velocidad media en relación con el suelo. Los círculos vacíos representan zonas en las cuales no llegaron a contemplarse aves terrestres en migración.

medio de todos los migradores en el cabo Cod, Wallops Island, Miami y Bermuda era hacia el sudeste, el análisis correspondiente a Antigua resultaba probablemente el más importante. Allí dividimos las aves en dos grupos: las que se desplazaban a lo largo de las Antillas hacia el sudeste y las que llegaban desde el Atlántico desplazándose hacia el sudoeste. A pesar de esta división arbitraria, el rumbo medio de ambos grupos era hacia el sudeste; la diferencia en la dirección de la ruta se debía al hecho de que las aves volaban bajo distintas condiciones del viento.

El análisis de los rumbos sugiere que para este vuelo de 3000 kilómetros es adecuado un sistema de orientación notablemente simple. Una vez las aves abandonan la costa de Norteamerica, mantienen un rumbo o marcación constante hacia el sudeste hasta que llegan a la costa de Sudamérica. Aparentemente, no es necesario que cambien este rumbo; el cambio de un curso sudeste a un curso sudoeste lo realizan cuando se encuentran con los vientos alisios del nordeste, en la zona del mar de los Sargazos. Una masa importante de pruebas experimentales apoya la idea de que un ave puede establecer y mantener un rumbo de derrota en base al sol, las estrellas o el campo magnético terrestre.

Aunque el radar ofrece muchas ventajas para el estudio de la migración de las aves, no nos permite identificar las aves que se detectan en su pantalla. Hemos abordado la cuestión de manera indirecta, determinando la velocidad en el aire de las 4600 o más aves que detectamos por radar. En las estaciones costeras la variación de las velocidades de vuelo es bastante amplia, pero en el Caribe éstas oscilan principalmente entre 30 y 60 kilómetros por hora. Si bien la velocidad de vuelo no constituye un método preciso para identificar a las aves, la mayoría de las que vuelan a menos de 45 kilómetros por hora son, probablemente, avecillas canoras. Las aves litorales y las acuáticas vuelan más raudas.

Se pueden identificar, con mayor exactitud, las aves que se ven pasar desde, o que se posan en, los barcos. Durante los dos últimos años de nuestras observaciones a bordo, Carol P. McClintock, de la Universidad estatal de Nueva York en Buffalo, se ocupó de obtener tantos ejemplares como fuera posible. Puesto que en los barcos no se permiten armas de fuego, se valió de trampas para hacerse con ejemplares vivos y de un tirachinas cuyas municiones consistían en uvas verdes. Sus hallazgos, junto a las cuidadosas observaciones realizadas con prismáticos, revelan que los pequeños paseriformes eran, con mucho, las especies más abundantes de las



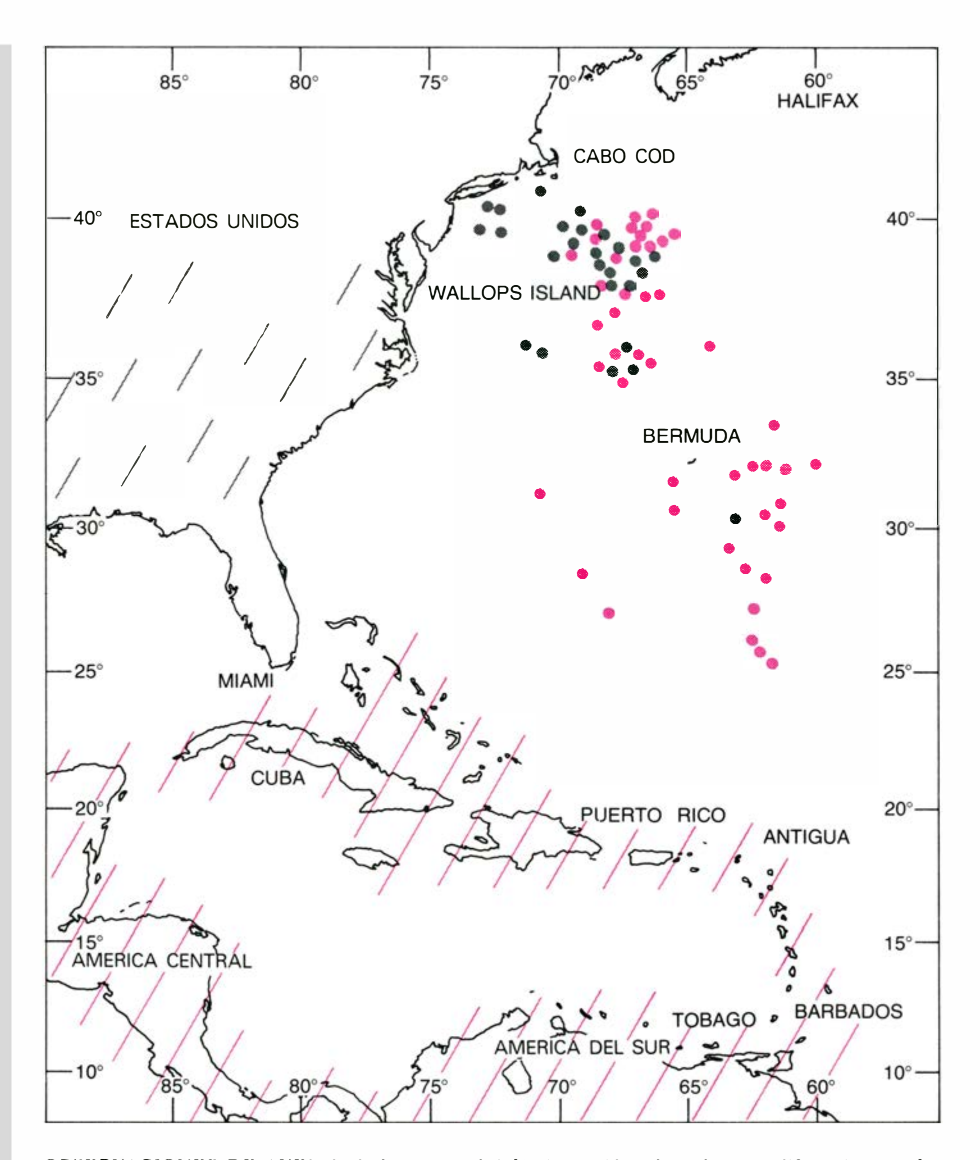
EFECTO DE LA METEOROLOGIA sobre el comportamiento migrador, cartografiando los datos de la ilustración precedente sobre un esquema de las condiciones meteorológicas típicas. Las estructuras meteorológicas son mucho más importantes que la geografía a la hora de influir sobre el comportamiento migrador.

avistadas desde barcos en todas las zonas del Atlántico que investigamos.

Las variaciones en la distribución de los pájaros observada parecía reflejar diferencias de comportamiento más que físicas. Por ejemplo, las currucas, los gorriones y los juncos tienen aproximadamente el mismo tamaño y el mismo peso y, sin embargo, todas nuestras observaciones de gorriones y juncos, menos una, correspondían a la zona situada entre la costa y Bermuda, mientras que se divisaron currucas tan al sur y al este cuanto los barcos se alejaron. Creemos que muchas aves que hemos definido como migradores verdaderos son currucas, mientras que otras muchas que se observan errando detrás de los sistemas frontales o al sur de las zonas de baja presión (los migradores desafortunados) pertenecen a las especies de gorriones y juncos que no migran al Caribe o a Sudamérica.

El Atlántico no es un lugar acogedor para los pájaros pequeños. Durante temporales en el mar se ha visto a cientos o a miles de avecillas volando alrededor de un barco, entrando en los camarotes y chocando con los mástiles o los cables. En estas ocasiones, no parecía posible que las aves sobrevivieran más allá de unas pocas horas. Nuestra opinión de que gran número de estas aves perecen en el mar se ve apoyada por los hallazgos de los oceanógrafos, quienes frecuentemente encuentran plumas de aves en el estómago de peces abisales del Atlántico.

Nuestros datos procedentes del radar, sin embargo, indican que estas pérdidas tienen lugar sobre todo en las aves que carecen de las adaptaciones del comportamiento que requiere el vuelo sobre el océano. Para lo migradores verdaderos, una estrategia migradora relativamente simple saca partido de una serie, asombrosamente predecible, de condiciones atmosféricas a lo largo de la ruta. En primer lugar, los migradores esperan en la costa oriental de Norteamérica hasta que un potente frente frío pasa sobre la costa y, en dirección sudeste, hacia el Atlántico. Esta estructura meteorológica no sólo asegura a las aves vientos favorables del noroeste durante la primera parte de su vuelo, sino que además es un indicador especialmente adecuado de un clima bueno entre la costa y el mar de los Sargazos. Las tormentas o los huracanes tropicales que se acercan a la costa norteamericana suelen frenar el movimiento hacia el sudeste de los frentes fríos. A lo largo de 93 días o noches observamos movimiento moderado o in-



OBSERVACIONES DE AVES desde barcos en el Atlántico occidental revelan una diferencia entre las pautas de migración de los pequeños pájaros que invernan en los Estados Unidos meridionales y los que invernan en el Caribe y en América del Sur. El primer grupo, representado por puntos grises y un rayado gris en la zona de invernada, comprende tres especies de gorriones y el junco pizarroso. El segundo grupo abarca ocho especies de picafollas y currucas americanas y está indicado por puntos y rayados de color. Algunas especies de currucas invernan en América Central y en América del Sur, pero no en las Antillas.

tenso de aves por la costa norteamericana o en Bermuda, y sólo en dos de estas ocasiones vimos aves que aparentemente volaban en el interior de la ruta de una tormenta que se acercaba.

Después de haber abandonado la costa, los migradores sólo tienen que volar constantemente hacia el sudeste hasta que alcanzan el Caribe o Sudamérica; su curso real será primero hacia el sudeste hasta el mar de los Sargazos, y después, cuando encuentran los vientos alisios, hacia el sudoeste, hasta alcanzar su destino. Tan pronto penetran en los sistemas frontales cercanos a Bermuda, su vuelo se desarrollará en su mayor parte bajo cielos despejados, con vientos suaves y variables o vientos moderados y constantes del nordeste. La ruta oceánica de Halifax a Antigua es unos 2800 kilómetros más corta (es decir, el 47 por ciento) que otra ruta que pase por Florida. Además, las aves que siguen la ruta oceánica aprovechan los vientos costeros del noroeste y los alisios del nordeste como vientos de cola, y la ruta presumiblemente carece de depredadores. Sin embargo, el viaje requiere un grado de esfuerzo que no tiene parangón en ningún otro vertebrado. Para el hombre, el equivalente metabólico sería correr durante 80 horas a una velocidad de un kilómetro cada 2 minutos y medio. Los sistemas respiratorio y muscular de las aves presentan muchas adaptaciones que hacen posible este rendimiento. Si un picafollas estriado quemara gasolina en lugar de sus reservas de grasa corporal, bien podría jactarse de obtener 300.000 kilómetros por litro de combustible.

Juegos matemáticos

Acertijos y problemas de teoría de números que nos suscitan las curiosas fracciones del antiguo Egipto

Martin Gardner

Mucho antes de comenzar la Era Cristiana, los matemáticos chinos tenían ya conocimientos sorprendentemente refinados sobre los quebrados. Admitían que el numerador o denominador pudiese ser un entero cualquiera, y disponían de excelentes reglas para sumar, restar, multiplicar y dividir fracciones. Lo mismo que hoy suele hacerse, los chinos preferían trabajar con fracciones propias, es decir, fracciones cuyo numerador es menor que su denominador, característica que se refleja en la nomenclatura: *tzu* (hijo), para el numerador, y *mu* (madre), para el denominador.

Los métodos de los antiguos egipcios eran, por el contrario, peculiarmente renqueantes. Comprendían bastante bien las fracciones de numerador mayor que 1, pero al parecer no sabían manejarlas como un único número. Con la única excepción de 2/3, para el cual crearon un jeroglífico especial, sólo tenían símbolos para fracciones unitarias, es decir, fracciones que sean recíprocsas de enteros positivos, con 1 situado sobre la raya de fracción, y un entero positivo cualquiera bajo ella.

Para manejar fracciones con numeradores mayores que la unidad, los egipcios descomponían tales fracciones en suma de varias fracciones unitarias. Por ejemplo, en lugar de escribir 5/6 escribirían 1/2 + + 1/3. Pusieron a punto las reglas precisas para poder efectuar las operaciones aritméticas necesarias sobre fracciones de este tipo. En ciertos casos, sobre todo cuando se trata de sumas, puede ser ventajoso trabajar con fracciones en forma desarrollada, pero en general, los métodos chinos son, con mucho, superiores.

La mayor parte de lo que sabemos acerca de las fracciones de los antiguos se ha deducido de la información aportada por un famoso documento hoy conocido por "papiro de Rhind", cuya inscripción debió realizarse hacia 1700 a. C. Comprado en Luxor, en 1858, por un anticuario escocés, A. Henry Rhind, en la actualidad es propiedad del British Museum. El papiro, una especie de manual de calculista, comienza por una tabla donde toda fracción de la forma $2/b$ se descompone en suma de varias fracciones unitarias ordenadas por valor decreciente, siendo b igual a cada uno de los enteros impares, desde 5 hasta 101.

¿Disponían los egipcios de un método sistemático para desarrollar de esta forma cualquier fracción propia? Cierto número de eruditos han especulado sobre tal posibilidad, pero lo más probable parece ser que no, ya que los desarrollos dados en el papiro de Rhind no siempre son "óptimos". Evidentemente, "óptimo" puede interpretarse de muchas formas. La más evidente sería considerar óptimos los desarrollos que hagan mínimo el número de sumandos. Otro tipo de desarrollo óptimo sería el que hiciera mínimo el máximo valor de los denominadores de las fracciones del desarrollo. Por ejemplo, al desarrollar 3/7 en la forma 1/4 + 1/7 + 1/28 se hace mínimo el número de sumandos, no obstante el desarrollo 1/6 + 1/7 + + 1/14 + 1/21 hace que el mayor de los denominadores tenga el valor mínimo posible. Si con un mismo desarrollo se consigue hacer mínimo el número de sumandos y también el denominador máximo, tanto mejor. (No entraré a considerar otros tipos de desarrollos óptimos, como podrían ser los que dieran mínimo valor a la suma de denominadores.)

Todavía más sorprendente que la afición de los egipcios por tan fastidioso método es que los griegos lo adoptaran tras ellos. De hecho, el sistema fue extensamente utilizado en Europa hasta bien entrado el siglo XVII. El propio Arquímedes usaba en sus cálculos las hoy denominadas fracciones egipcias. Esta denominación ha venido a designar cualquier desarrollo de un número racional en suma de varias fracciones unitarias, que es costumbre ordenar por denominadores crecientes. Las fracciones modernas, en las que pueden encontrarse por encima y debajo de la línea enteros cualesquiera, proceden de la matemática hindú, y no se adoptaron de forma generalizada hasta el siglo XVIII. Algunos historiadores han mantenido que la prolongada obsesión por las fracciones unitarias ha sido una tendencia cultural que retrasó el avance de las matemáticas, casi tanto como el engorroso sistema de numeración romana.

En teoría de números, el estudio e investigación de las fracciones egipcias es hoy tarea de poca monta, aunque nada sencilla. Hay en este terreno muchas y profundas cuestiones todavía no resueltas; aunque también hay otras muchas al alcance de novicios despiertos, muy relacionadas con ciertas curiosidades y problemas de matemática recreativa. Tomemos, por ejemplo, el viejo rompecabezas de origen árabe sobre un hombre que impuso en su testamento la condición de que sus 11 caballos fuesen repartidos entre sus hijos de forma que el mayor recibiera 1/2, el mediano 1/4, y el menor de los hermanos, 1/6 de los caballos. Al morir, sus albaceas quedaron perplejos, no sabiendo cómo ejecutar tan excéntricas instrucciones. Después de todo, de poco vale un caballo hecho rodajas. Un pariente, enterado del problema, lo resolvió prestándoles su propio caballo. Fue fácil entonces repartir los caballos según lo convenido en el testamento, recibiendo los hijos seis, tres y dos caballos, respectivamente. Sobraba todavía un caballo, con lo que el pariente pudo también recobrar el suyo...

Este problema se propone bajo multitud de formas, y desde luego, puede generalizarse fácilmente, manejándose mayor número de hermanos y caballos prestados y más tarde recuperados. Ciñéndonos a la forma tradicional del cuento, donde hay tres hijos y solamente se presta un caballo, se puede plantear una interesante cuestión: ¿Cuántas variantes hay para el número de caballos a repartir y para el conjunto de tres fracciones especificadas en el testamento del padre? Podríamos conjeturar que tal vez infinitas; resulta, sin embargo, que solamente hay siete. Son las siete soluciones de la ecuación diofántica $n/(n + 1) = 1/a + 1/b + 1/c$, siendo a, b y c enteros positivos diferen-

	n	a	b	c
1	7	2	4	8
2	11	2	4	6
3	11	2	3	12
4	17	2	3	9
5	19	2	4	5
6	23	2	3	8
7	41	2	3	7

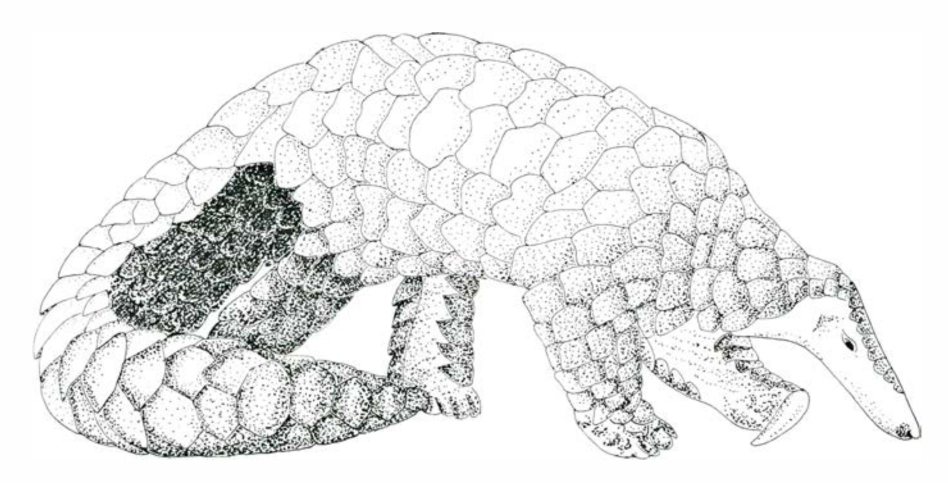
Todas las variantes de un antiguo acertijo árabe

tes, y además, tales que a es menor que b, b es menor que c y $n + 1$ es el mínimo común múltiplo de a, b y c.

Es fácil demostrar que a debe ser igual a 2. Si a fuese mayor que 2, el mínimo común múltiplo de a, b y c más bajo posible sería 12, y se obtendría cuando a fuera igual a 3, b igual a 4 y c igual a 6. Así pues, $n/(n + 1)$ tiene que ser al menos 11/12. Pero la suma 1/3 + 1/4 + 1/6 es igual a 47/60, que es menor que 11/12, y si se aumentan los denominadores, la suma resultante es todavía más pequeña. Por consiguiente, a es menor o igual que 2, y, así pues, tiene que ser igual a 2. Un razonamiento parecido muestra que b tiene que ser, o bien 3, o bien 4, y con esta información ya no es difícil determinar todos los valores posibles de c. La tabla de la ilustración de la página opuesta da el valor de n (número inicial de caballos a repartir) y los denominadores de las tres fracciones unitarias correspondientes a las siete variantes del problema.

Es evidente que toda fracción propia puede descomponerse en suma de fracciones unitarias, si se permite repetir sumandos. Por ejemplo, 3/7 es igual a 1/7 + + 1/7 + 1/7. No es obvio, sin embargo, que todas las fracciones propias puedan descomponerse en suma de fracciones unitarias todas distintas. Demostración de que así ocurre es que existe un famoso algoritmo que permite expresar cualquier fracción propia en suma de un número finito de fracciones egipcias distintas. El algoritmo fue publicado por primera vez en 1202, por Leonardo de Pisa, más conocido por el apodo de Fibonacci, en un libro que ejerció gran influencia: el *Liber abaci*. Fibonacci prefería trabajar con fracciones unitarias, y su libro contiene tablas de conversión de fracciones propias en suma de fracciones egipcias. Su algoritmo para traducir fracciones propias en suma de egipcias diferentes se presenta sin demostración alguna de que vaya a funcionar siempre. El eminente matemático británico J. J. Sylvester redescubrió el algoritmo, y en 1880 publicó la primera demostración de la validez general del método.

El método de Fibonacci es sencillo. Llamemos a/b a la fracción propia dada. El primer sumando del desarrollo es la máxima fracción unitaria no mayor que a/b. Ahora se substrae de a/b esta fracción unitaria, con lo que resulta una nueva fracción propia. La segunda fracción del desarrollo es la máxima fracción unitaria no mayor que esta diferencia. Se prosigue de esta forma, extrayendo cada vez la máxima fracción unitaria utilizable, que será el siguiente término del desarrollo, restando, y repitiendo el proceso sobre la diferencia. Es evidente que de esta forma las fracciones obtenidas irán decreciendo progresivamente. Puede demostrarse que el proceso siempre concluye en un número finito de pasos. Por tanto, el algoritmo funciona siempre. (También es posible expresar cualquier número irracional como suma de fracciones egipcias diferentes, que formarán ahora una serie infinita, aunque esta cuestión se sale demasiado de nuestro tema.) En terminología moderna, el método de Fibonacci es un algoritmo voraz, porque en cada etapa del proceso elige la fracción mayor posible.



Pangolino "desenrollado"

Aunque el algoritmo voraz permite expresar cualquier fracción propia a la manera egipcia, no siempre los desarrollos obtenidos son óptimos en ninguno de los sentidos que antes dimos a este calificativo. No obstante, aplicado a fracciones de la forma $1/b$ sí genera un desarrollo binómico de $1/b$, que es óptimo en ambos sentidos. Un poco de cálculo algebraico bastará para convencer al lector de que aplicando el algoritmo a $1/b$ resulta $1/(b + 1) + 1/[b(b + 1)]$. Así, 1/2 es igual a 1/3 + 1/6; 1/3 es igual a 1/4 + 1/12; 1/4 es igual a 1/5 + 1/20, y así sucesivamente.

La fórmula anterior, que da un desarrollo binómico de $1/b$ en suma de fracciones egipcias, puede servir también para demostrar que toda fracción propia puede expresarse al modo egipcio de infinidad de formas. Tomemos el desarrollo 2/3 = = 1/2 + 1/6. Aplicando la fórmula al último sumando, 1/6, tenemos un nuevo desarrollo: 2/3 = 1/2 + 1/7 + 1/42. Iterando el proceso con 1/42 tenemos 2/3 = 1/2 + 1/7 + 1/43 + 1/1806. Vemos así que el desarrollo de 2/3 como suma de fracciones unitarias puede proseguir indefinidamente.

Esta misma fórmula inspira otro algoritmo, llamado método de descomposición que, al igual que el algoritmo de Fibonacci, garantiza un desarrollo en serie finita egipcia para toda fracción propia. Hay otros muchos algoritmos que cumplen igual propósito, cada uno con sus propios méritos y deméritos. Algunos hacen mínimo el número del sumando; otros, el denominador máximo, pero todos, incluido el algoritmo voraz, son ineficientes, y de incómoda aplicación a fracciones cuyo numerador y denominador sean grandes.

Al aplicar el algoritmo voraz a una fracción propia a/b se genera una serie de fracciones egipcias cuyo número de términos es siempre menor o igual que a. Por tanto, aplicado a fracciones de la forma $2/b$, genera una expresión con uno o dos sumandos: 2/4 igual a 1/2; 2/5 igual a 1/3 + 1/15; 2/6 igual a 1/3; 2/7 igual a 1/4 + 1/28, y así sucesivamente. En cada paso el algoritmo elige la máxima fracción unitaria que es menor que la restante, pero al no ser afectado ningún paso por los resultados ni de los precedentes ni de los sucesivos, al aplicar este procedimiento a fracciones propias con numerador 3 o mayor es fácil pasar por alto otros desarrollos más económicos. Además, este algoritmo propende a generar términos cuyo denominador es mucho mayor de lo necesario.

Michael N. Bleicher, en su sección sobre fracciones egipcias del libro *Excursions into Mathematics* (Worth Publishers, 1969) presenta varios aterradores ejemplos que muestran cuán abyectamente puede fracasar el algoritmo voraz en lo que respecta a obtener desarrollos óptimos, en cualquiera de los dos sentidos del término. Por ejemplo, al aplicar este algoritmo a la fracción 5/121 se obtiene la siguiente serie5/121 = 1/25 + 1/757 + + 1/763.308 + 1/873.960.180.913 + + 1/7.638.092.437.828.241.151.744. Bleicher compara la serie anterior con el desarrollo 5/1/21 = 1/25 + 1/759 + + 1/208.725, que es mínimo. No es posible descomponer 5/121 en menos de tres sumandos, pero Bleicheer ignora si existe o no un desarrollo de tres términos cuyo máximo denominador sea inferior a 208.725.

Para fracciones propias de la forma $3/b$ el algoritmo voraz garantiza desarrollos de a lo máximo tres sumandos; si la frac-

ción fuese de la forma $4/b$, el desarrollo correspondiente no tendría más de cuatro. Una notable cuestión sobre fracciones egipcias, todavía no resuelta, se refiere al caso $4/b$: ¿Se puede desarrollar siempre una fracción propia $4/b$ sin usar más de tres términos? Dicho de otra forma, ¿se puede resolver siempre la ecuación diofántica $4/n = 1/a + 1/b + 1/c$, cualquiera que sea el valor entero n, mayor que 4?

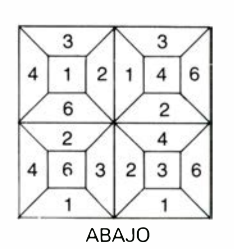
Paul Erdös y E. G. Straus han enunciado la conjetura de que la ecuación anterior sí puede resolverse siempre. Su conjetura ha sido verificada hasta valores extraordinariamente grandes de n, pero todavía no ha podido ser demostrada. Análogamente, Waclaw Sierpinski ha expuesto la también conjetura de que todas las fracciones de la forma $5/b$ pueden descomponerse en no más de tres sumandos egipcios. Más todavía, ha conjeturado asimismo que dado un entero fijo k, arbitrario, y una variable b de valores enteros, existe un valor de b, mayor que k, tal que, para todo valor de b mayor que el anterior, la fracción k/b puede descomponerse siempre con no más de tres términos.

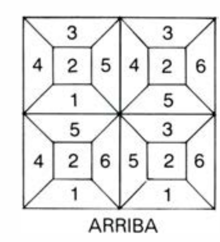
Hasta el momento solamente hemos considerado fracciones propias. ¿Qué podrá decirse sobre fracciones impropias, tales como 7/3 o 2/1? También estas fracciones pueden desarrollarse siempre en serie egipcia, y esto, de infinitas formas. Para engendrar su desarrollo comenzaremos por los términos de la serie armónica, $1/1 + 1/2 + 1/3 + 1/4...$, por ser éstas las máximas fracciones unitarias que podemos utilizar. Es bien sabido que la serie armónica no es convergente. Dicho de otra forma: no importa qué número entero podamos tomar; siempre habrá una suma parcial de la serie (suma de los n primeros términos, para algún valor de n) que lo supere. El procedimiento para generar una serie de fracciones egipcias que sea igual a una fracción impropia dada consiste en utilizar sumas parciales de la serie armónica mientras sea posible, y añadir después otras fracciones unitarias que expresen la parte fraccionaria que pudiera restar para alcanzar el valor requerido. Sin embargo, la serie armónica diverge con lentitud exasperante, cada vez más despacio, y por este motivo, incluso pequeñas fracciones impropias suelen requerir desarrollos egipcios enormemente largos. Por ejemplo, el desarrollo de 10/1 exige utilizar más de 20.000 fracciones unitarias de la serie armónica.

Se han obtenido recientemente algunos curiosos resultados sobre fracciones egipcias. En 1963 se demostró que todo entero positivo puede expresarse mediante una serie egipcia cuyos denominadores formen progresión aritmética. En 1964, Ronald L. Graham, de los Bell Laboratories, estudió el problema de hallar qué fracciones racionales pueden ser expresadas mediante fracciones egipcias cuyos denominadores sean cuadrados perfectos. No se contentó con resolver este problema por completo, sino que además solventó el caso general, a saber, determinar qué fracciones pueden descomponerse en fracciones egipcias cuyos denominadores sean todos potencias de un mismo grado mayor que 2. Ese mismo año, Graham demostró también que, dado un conjunto que contenga todos los números primos mayores que cierto número, así como todos los cuadrados perfectos mayores que cierto número (posiblemente distinto del anterior), entonces toda fracción racional admite un desarrollo egipcio cuyos denominadores proceden todos de dicho conjunto.

Al prescribir que todos los denominadores de las fracciones egipcias utilizadas sean números impares surgen problemas de especial dificultad. Es fácil ver que la suma de una serie de tales fracciones no pueden ser una fracción irreducible con denominador par. Se ha demostrado, por otra parte, que toda fracción racional con denominador impar es descomponible en serie de fracciones egipcias distintas, todas con denominadores impares, como, por ejemplo, en $2/3 = 1/3 + 1/5 + 1/9 + 1/45$; $2/5 = 1/3 + 1/15$, y $2/7 = 1/7 + 1/9 + 1/35 + 1/315$. Existen algoritmos ineficientes para hallar tales desarrollos, pero nadie ha podido demostrar que el algoritmo voraz de Fibonacci finalice siempre en un número finito de pasos al aplicarlo a esta tarea, ni siquiera para fracciones propias. Admitiendo que los denominadores puedan ser pares o impares, el algoritmo voraz genera fracciones unitarias progresivamente menores, y puede probarse que la serie ha de concluir. Sin embargo, al exigir que los denominadores sean forzosamente impares, las fracciones de la serie que genera el algoritmo crecen y decrecen, al parecer aleatoriamente, y la tarea de probar que la serie finaliza es formidable.

El desarrollo de 1 mediante el mínimo número de fracciones egipcias con denominadores impares no fue descubierto hasta 1976. (La expresión 1/1 no es, evidentemente, lícita.) Resulta que este problema tiene cinco soluciones, todas formadas por nueve sumandos. Entre ellas, la de mínimo denominador máximo es $1 = 1/3 + 1/5 + 1/7 + 1/9 + 1/11 + 1/15 + 1/35 + 1/45 + 1/231$. Los cinco desarrollos mencionados comienzan siempre por las recíprocas de 3, 5, 7, 9, 11 y 15. Las cuatro soluciones restantes prosiguen con los recíprocos de los números 21, 135 y 10.395; 21, 165 y 693; 21, 231 y 315; y 33, 45 y 385. ¿Cuál de los





Solución de "La locura en ocho bloques"

desarrollos de 1 con todos los denominadores impares tiene mínimo denominador máximo? La solución, única, está dada por una serie de 11 términos: $1 = 1/3 + 1/5 + 1/7 + 1/9 + 1/11 + 1/33 + 1/35 + 1/45 + 1/55 + 1/77 + 1/105$.

Sumando a cada una de las series anteriores la fracción 1/1, se obtienen descomposiciones óptimas entre las de denominadores impares para el número 2. Sin embargo, ignoro cuáles sean los desarrollos óptimos de 2 (en cualquier sentido del término) con denominadores impares si se prohíbe utilizar 1/1, y ni siquiera sé si tales desarrollos han sido descubiertos.

He aquí cuatro sencillos problemas de fracciones egipcias cuyas soluciones se darán el mes próximo:

1. Expresar 1 en suma de tres fracciones unitarias distintas.

2. Expresar 67/120 como suma del mínimo número de fracciones egipcias posible, con la condición de que el máximo denominador correspondiente a ese número de términos sea mínimo.

3. La fracción 8/11 es la "mínima" fracción propia que no puede expresarse con menos de cuatro fracciones egipcias, en el sentido de que hace mínima la suma de su numerador y su denominador. Hallar un desarrollo de esta fracción con cuatro sumandos.

4. Al aplicar el algoritmo voraz a una fracción propia de la forma $3/b$, ¿cuál es el mínimo valor de b tal que el algoritmo produce un desarrollo trinomio y que sin embargo $3/b$ puede expresarse como suma de solo dos fracciones unitarias distintas?

Al discurrir sobre animales matemáticamente interesantes en la sección de agosto cometí dos errores, uno por ignorancia, otro por descuido. Allí escribí que ningún animal avanza por el suelo rodando como un disco o una esfera. Desde luego, esto no es estrictamente cierto si se admite la posibilidad de rodar cuesta abajo. Numerosos lectores, me han escrito para decirme que, existe un pequeño arácnido que habita en madrigueras excavadas en las laderas de las dunas arenosas. Cuando este arácnido es atacado por una avispa, extiende las patas como si fueran los radios de una rueda y huye rodando

ladera abajo. Peter G. Trei, de Bélgica, me envió copia de una nota aparecida en el *Journal of Mammalogy* (febrero, 1975), enviada por un zoólogo americano, Richard R. Tenaza. Tenaza explica de qué forma eluden los pangolinos de la isla de Siberut, al oeste de Sumatra, ser capturados. Los pangolinos, especie de osos hormigueros recubiertos de escamas, se hacen un ovillo y se dejan rodar rápidamente cuesta abajo. De hecho, el mismo nombre "pangolino" procede de una palabra malaya que significa "rodar". Mi descuido consistió en situar incorrectamente la orca (pájaro imaginario cuya cola helicoidal le sirve para propulsarse) en un libro de la serie del mago de Oz. La orca pertenece al titulado *The Scarecrow of Oz* .

Chandler Davis amplió mi relación de animales imaginarios que ruedan haciéndome notar la existencia de una tal creatura en la fantasía de George MacDonald titulada *The Princess and Curdie.* Rufus P. Isaacs, comentando el hecho que ninguna esfera puede ser teselada con hexágonos, me proporcionó la demostración de un teorema no menos sorprendente que descubrió hace ya muchos años: Si se tesela una esfera con pentágonos y hexágonos, entonces tendrá que haber exactamente 12 pentágonos, ni más ni menos.

El primero de los problemas del mes pasado consistía en usar ocho cubos de color especificados para construir con ellos un cubo mayor, donde cada cara fuese de un único color sólido, y no hubiera dos caras del mismo color. Se muestran en la ilustración de la página opuesta las capas superior e inferior de una solución. La segunda solución se obtiene desplazando los cubos de la capa superior de forma que cada cubo pase a ocupar la posición adyacente en sentido contrario a las agujas del reloj, sin modificar para nada su orientación. El prototipo del modelo es el cubo *Fc* de la matriz de 30 cubos de color que se dio el mes pasado.

El segundo problema consistía en utilizar los mismos ocho cubos para construir un cubo mayor que tuviera los cuatro colores diferentes en cada cara, estando cada color representado seis veces en total. Se obtienen dos soluciones a partir de las dos del problema precedente. Basta con "trisecar" cada cubo, intercambiando las rebanadas izquierda y derecha, las rebanadas frontal y trasera y las rebanadas superior e inferior.

Las soluciones del problema de los 27 dados que aparecen en el número de octubre están orientadas de forma incorrecta. Para corregir los diagramas debe dársele a cada dado un cuarto de vuelta hacia arriba, y después deben intercambiarse las capas superior e inferior.

Taller y laboratorio

Algunas "salas de los secretos" se distinguen por reflejar el sonido, pero hay otras más intrigantes

Jearl Walker

Las "salas de los secretos", que suelen ser estancias abovedadas, constituyen una atracción para sus visitantes en cualquier parte del mundo. La característica que las distingue de las demás salas es que en ellas cualquier sonido, aunque sea tan débil como un murmullo, recorre distancias muy superiores a las que dicho sonido recorrería sin la ayuda acústica que le presta la estructura de la estancia. De pronto, el oyente encuentra el efecto sorprendente, en particular cuando está viendo la distancia a que se halla la persona que está susurrando lo que él está oyendo, pareciéndole imposible que el susurro pueda recorrer dicha distancia.

El efecto acontece principalmente con dos tipos de estructuras. En el tipo más simple, la superficie interior de la habitación tiene la forma de sección de una esfera o de un elipsoide. Cuando alguien se pone a hablar en el centro de curvatura de la esfera, las ondas sonoras emitidas se reflejan y vuelven a concentrarse hacia el individuo que habla, con sorprendente sonoridad. Si se trata de un elipsoide y el sujeto se sitúa en uno de los focos mientras que el oyente se coloca en el otro, igual que antes la superficie curva refleja las ondas sonoras emitidas por el hablante, concentrándolas en el oyente, en esta ocasión situado en el segundo punto focal. En ambos tipos de estructura, un sonido puede cruzar un gran volumen de aire y seguir siendo audible debido a la reflexión simple y a la concentración en los focos.

El segundo tipo de "sala de los secretos" tiene también las paredes curvadas, pero su análisis implica una mayor dificultad, pues no exige la concentración del sonido en los focos. Cuando una persona emite un murmullo a lo largo de una de estas paredes circulares, el sonido se mantiene, de alguna manera, en una capa adyacente a la pared; se propaga a lo largo de la misma y puede ser oído por otra persona situada en cualquier punto de la circunferencia. Por qué permanece el sonido pegado a la pared y cómo puede el lector efectuar algunas demostraciones de este efecto son el tema de mi trabajo de este mes, aunque quiero dejar sentado, desde un principio, que no tengo claro todo lo referente a este tema.

La más famosa "sala de los secretos" de este segundo tipo está en la base de la cúpula de la catedral de St. Paul de Londres. La estancia es un corredor situado a unos 40 metros sobre el nivel del piso principal. Tiene una anchura de 1,8 metros y forma un círculo con un diámetro de 3,3 metros. La pared lateral a lo largo del corredor tiene una ligera inclinación hacia el centro. El corredor dista del vértice de la cúpula unos cuarenta metros.

Desde hace tiempo se sabía que si alguien murmuraba cualquier cosa a lo largo de la pared de la sala, dicho sonido podía oírse por otra persona que estuviese situada en cualquier punto de la circunferencia del corredor, siempre que se pegara a la pared. Por el contrario, si el bisbiseo se emitía hacia el oyente, pero siguiendo una dirección diametral, en el círculo de la estancia, las ondas sonoras se esparcían y el sonido se hacía inaudible para este último.

La anormal acústica de la sala llamó la atención científica, en primer lugar, de Sir John Herschel, reputado sabio que se le reconoce por sus contribuciones a la óptica y a la astronomía en la primera mitad del siglo XIX. Más tarde (en 1871), Sir George Airy, astrónomo real, explicó el efecto en términos de simple reflexión y concentración de las ondas sonoras por la cúpula. Sin embargo, en 1878, Lord Rayleigh estableció que tal explicación podría no ser correcta, ya que, para que así ocurriera, el oyente debería hallarse en el punto del corredor diametralmente opuesto al de la persona que hablaba. Puesto que el oyente podía situarse realmente en cualquier otro punto del pasillo circular, era inmediato que el efecto no se debía a las reflexiones de la bóveda.

La primera explicación de Rayleigh apareció en su *Theory of Sound* ("Teoría del sonido") (volumen II, sección 287, publicado en 1878). En ella expuso sus últimos trabajos sobre la difracción del sonido al salir de la boca de una persona. En general, cuando las ondas sonoras atraviesan una apertura de una anchura comparable a su longitud de onda, se interfieren una con otra y dan lugar a que el conjunto sufra una difracción. El sonido más intenso se encuentra en el centro; a partir del centro, y formando ángulo hacia fuera, se encuentran alternativamente regiones de interferencia destructiva (correspondiente a menos intensidad de sonido) y de interferencia constructiva (correspondiente a un sonido más intenso, aunque no tanto como en la región central). Acontece una difracción similar cuando las ondas sonoras pasan por un borde o se mueven alrededor de un objeto que tenga un tamaño equiparable al de su longitud de onda.

Cuando una persona habla, el sonido sale de su boca con una determinada configuración ondulatoria debido a que el sonido se difracta primeramente al cruzar la abertura de la boca y después se dirige hacia atrás, rodeando la cabeza. Es más difícil determinar dicha forma de lo que sería si el sonido saliera por una simple abertura de cualquier otra superficie sólida; por ello, y de momento, podemos simplificar la forma concentrándonos en el pico central de intensidad en un experimento ideal realizado con una simple apertura. La amplitud angular de la región central depende, en parte, de la longitud de onda del sonido; cuanto mayor sea esta longitud de onda (o lo que es lo mismo, más baja sea la frecuencia) tanto más ancha será la región central. Si un observador se pone a escuchar a una persona que esté emitiendo sonidos, principalmente de corta longitud de onda (esto es, de alta frecuencia), aunque se halle frente a ella es posible que no esté dentro de la zona central de las ondas difractadas. Pero si el sonido emitido está fundamentalmente en la zona de las longitudes de onda mayores, el observador oirá el sonido con nitidez aun cuando no se encontrara ante la persona que lo está emitiendo, ya que la difracción hace que la región central sea lo suficientemente ancha como para que se envíe una significativa cantidad de sonido hacia los lados del que habla e incluso hacia atrás del mismo.

Rayleigh estableció que, para que el corredor de la catedral de St. Paul se comportara como una "sala de los secretos", era necesario que las personas susurrasen a lo largo de la pared. Por ser los susurros sonidos de alta frecuencia, la mayoría emergen como un cono muy estrecho a lo largo de la pared. Por mor de simplificación consideremos que el cono sólo tenga dos dimensiones, en lugar de tres, y que se extienda en el plano horizontal (prescindiremos, por tanto, de la propagación vertical del sonido).

El sonido se separa de la pared formando un cono de rayos con un ángulo máximo θ con la tangente a dicha pared.

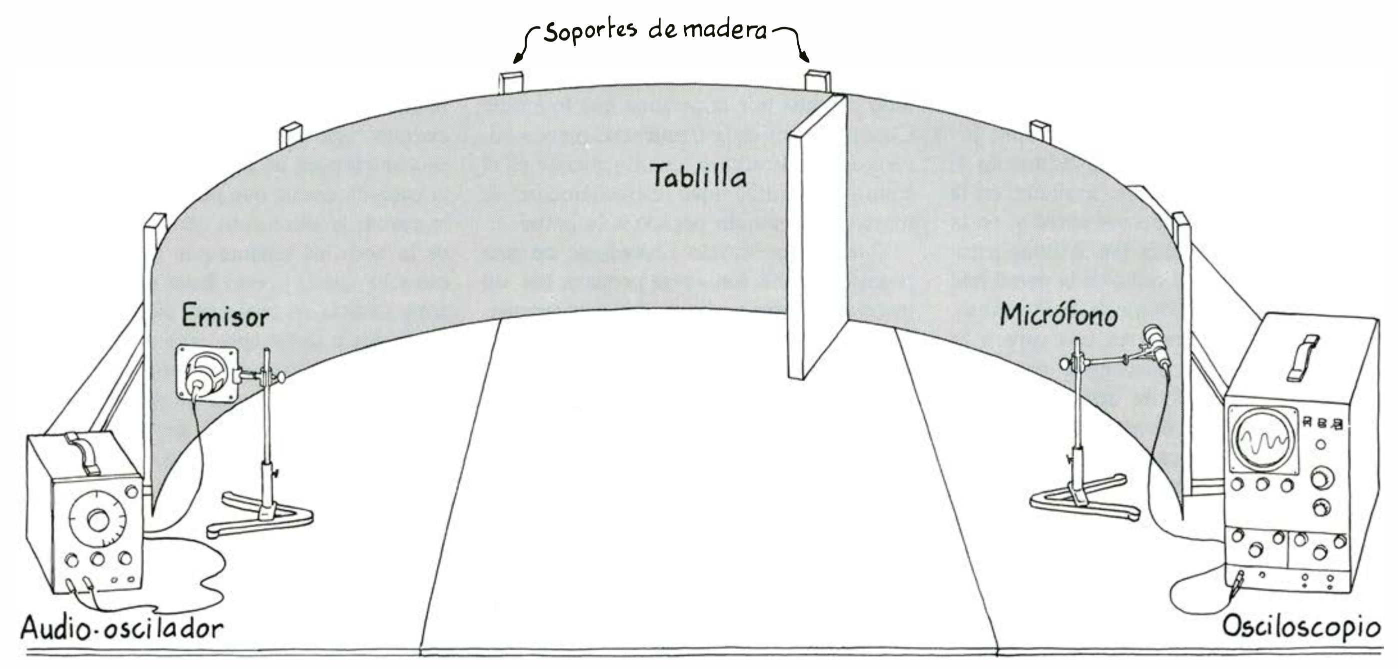
Los rayos se reflejan sobre la pared curvada en diversos puntos, de tal modo que en cada reflexión el ángulo de reflexión es igual al ángulo de incidencia. Como los rayos comienzan en un cono de ángulo θ, permanecen en una región limitada, en la parte exterior, por la propia pared y, en la parte interior, por una pared imaginaria con un radio igual al radio de la pared real multiplicado por el coseno de θ. Por tanto, los rayos se mantienen en una capa a lo largo de la pared y pueden ser oídos por cualquiera que se sitúe dentro de dicha zona, pero no fuera de ella. En este ejemplo el espesor de la capa está determinado por el radio de curvatura de la galería y por el ángulo máximo con el cual el sonido se separa de la pared cuando alguien susurra a lo largo de ella. Este ángulo viene determinado a su vez por la difracción del sonido causada por la persona que lo emite. Cuanto mayor es la frecuencia, menos notoria es la difracción; cuanto menor es el ángulo máximo, más estrechamente se mantiene el sonido pegado a la pared.

Cuando un sonido procedente de una pequeña fuente sonora se propaga por un plano en el espacio libre, decae su intensidad de un modo inversamente proporcional al cuadrado de la distancia recorrida. Sin embargo, con las reflexiones múltiples —en el modelo de Rayleigh para la galería de St. Paul— la intensidad disminuye proporcionalmente a tan sólo el inverso de la primera potencia de la distancia recorrida. Así pues, el sonido permanece con mayor vigor al recorrer la pared, que si se hubiera propagado siguiendo el diámetro de la circunferencia. Los sonidos de frecuencia más baja, presentes en la voz humana, no contribuyen de un modo significativo a la capa de ondas que permanece pegada a la pared; la difracción que sufren al salir de la boca las esparce por fuera de dicha estrecha capa, y esto hace que otra persona situada en cualquier punto de la circunferencia intercepte sólo una pequeña cantidad de la intensidad original. El modelo de Rayleigh gozó de gran prestigio debido a su simplicidad, pero como veremos más adelante dicha simplicidad puede conducir a errores.

En 1904, Rayleigh volvió sobre el efecto de la "sala de los secretos", traba-



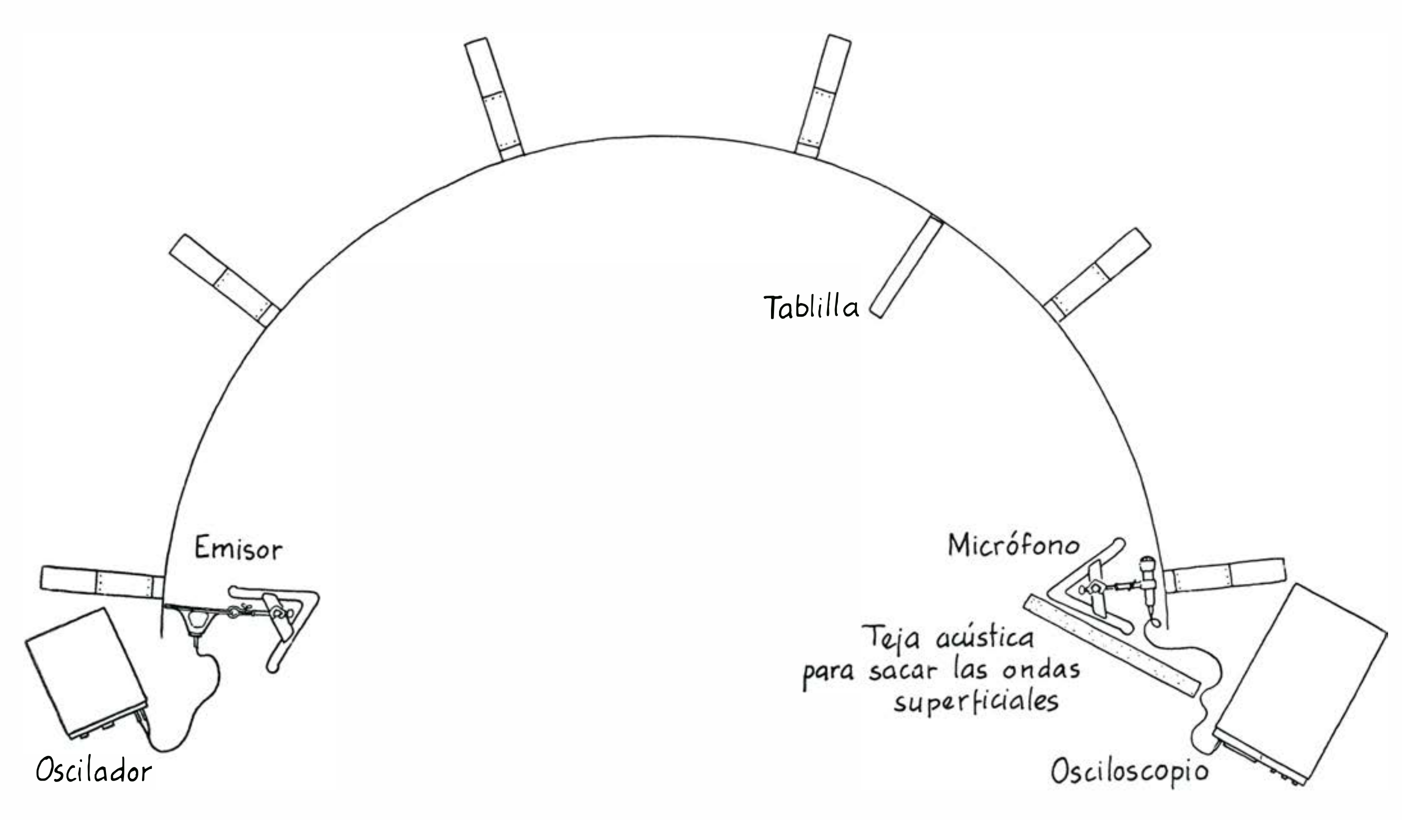
"Sala de los secretos" de la Catedral de St. Paul en Londres **(debajo de las ventanas superiores)**



Montaje de John S. Derov para comprobar el efecto de St. Paul

jando con un modelo de laboratorio consistente en una gran lámina de cinc (3,6 × 0,6 metros) doblada en forma de semicírculo. En el lado interior de uno de los extremos colocó un silbato, de los usados para reclamo de los pájaros, y en el otro extremo una vela que era sensible a las variaciones de presión producidas por el sonido. Al soplar, el silbato emitía un sonido de alta frecuencia con una longitud de onda de unos dos centímetros, provocando la oscilación de la vela, situada en el extremo opuesto. Pero cuando Rayleigh colocaba una tablilla estrecha (con una anchura de unos 5 centímetros) en cualquier sitio de la cara interior de la lámina de cinc, la vela no sufría ninguna perturbación imputable a los sonidos del silbato. Este resultado indicó a Rayleigh que las ondas sonoras se desplazaban a lo largo de la cara interior de la superficie curvada de la lámina de cinc en lugar de cruzar diametralmente el semicírculo. De forma semejante a lo que ocurría en la "sala de los secretos" de St. Paul, aquí también las ondas sonoras se desplazaban a lo largo de la pared.

A pesar del éxito de su modelo, Rayleigh no quedó satisfecho del todo. Y así, en 1910, publicó otro nuevo para explicar el efecto. En él estableció una correlación entre la oscilación del aire rodeado por la galería circular de la catedral de San Pablo y la oscilación de la membrana circular de un tambor. En un caso ideal, la membrana oscilaría con un tipo de onda con



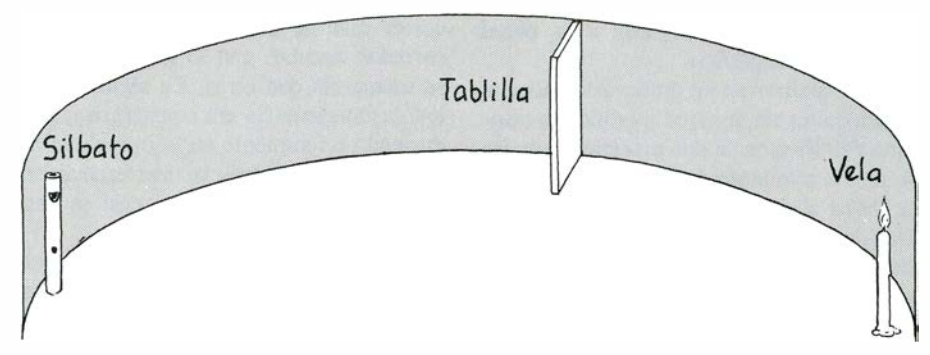
Vista en planta del montaje de Derov

amplitud nula a lo largo del borde (puesto que la membrana está sujeta en dicho lugar) y las amplitudes irían cambiando en dirección al centro de la membrana, de acuerdo con una ecuación en la que aparecen las funciones de Bessel, conocidas también como funciones cilíndricas (en razón del tipo de problemas que resuelven).

En una sala circular, una voz o cualquier otro ruido hace que las moléculas de aire oscilen, provocando, a su vez, la oscilación de la presión del aire. Estas oscilaciones de la presión del aire constituyen lo que nosotros llamamos sonido, porque hacen que nuestro tímpano vibre. Las moléculas próximas a la pared no pueden moverse (ya que encuentran la pared en su camino), pero las que están más lejos sí lo pueden hacer. Cuando estas moléculas más alejadas se mueven en dirección a la pared durante su oscilación, aprisionan a las moléculas más cercanas a la misma, aumentando la presión del aire. Cuando las moléculas más alejadas se mueven distanciándose de la pared, se separan de las moléculas que están más próximas a dicha pared, rebajando la densidad y, como consecuencia, disminuyendo la presión. Por eso las variaciones de presión en la capa próxima a la pared son relativamente grandes.

Para estudiar las características del movimiento de una membrana al vibrar en un tambor, la condición matemática que tiene que cumplirse es que, en el borde del tambor, la amplitud de las vibraciones sea nula. Al estudiar las variaciones de la presión del aire la condición a satisfacer es que, en la pared, las variaciones de la presión del aire alcancen un máximo. Para el oyente, eso significa que el sonido tendrá un máximo de intensidad en la pared. Conforme el oyente se vaya alejando de dicha pared, las variaciones de la presión y, por tanto, la intensidad del sonido irán disminuyendo de acuerdo con una función de Bessel, cuya variable sea la distancia del oyente al centro de la sala.

Para un tambor real y para una galería circular también real se pueden dar simultáneamente muchos modelos de éstos (correspondientes a las distintas funciones de Bessel resolutorias). Con el fin de simplificar, Rayleigh consideró que las oscilaciones en la galería se debían a una única función de Bessel de orden alto; este orden alto era necesario, pues las oscilaciones se daban sólo cerca de la pared y su intensidad se apagaba rápidamente al moverse hacia el centro del círculo. (Las funciones de Bessel de orden inferior producirían oscilaciones en zonas próximas al centro.) Tal excitación sólo se requería a lo largo de la pared, porque el individuo hablante susurraba cerca de la misma; consiguientemente, creaba una excitación mínima, sino nula, lejos de esa superficie.



Montaje experimental de Lord Rayleigh

Rayleigh estaba, por tanto, en condiciones de demostrar que las variaciones de presión provocadas por las ondas superficiales tenían su máximo en la pared y eran prácticamente nulas fuera de ellas. Además, estaba en condiciones de escribir una ecuación que diera la amplitud de la variación de la presión en función de la distancia al centro de la galería. Pero, una vez más, la simplificación de sus resultados puede inducir a error.

John S. Derov, estudiante de física en la Universidad estatal de Cleveland, diseñó una serie de experimentos sobre las ondas superficiales en una superficie curva similar a la de Rayleigh. Los resultados no pueden ser explicados de forma tan sencilla como los del autor inglés. Lo que significa que el lector tiene ante sí un sugestivo trabajo si es que el tema le interesa.

Derov fabricó una pared semicircular uniendo dos láminas de metal de 2,4 metros de largo por 60 centímetros de ancho con lo que obtuvo una sola pieza de 4,5 metros de longitud. La parte en que las láminas se solapaban la unió con remaches. El espesor del material de las láminas no era demasiado importante, con la salvedad de que, con un espesor mayor, la estructura sería más robusta. Para conseguir que la lámina completa se mantuviera formando un semicírculo encima de una mesa lo suficientemente grande, Derov colocó ocho soportes de madera por detrás de la lámina separados entre sí unos 60 centímetros. Para aumentar la masa y la rigidez de la pared, pegó a la cara exterior de la misma láminas de cemento e incluso grandes trozos de Play–Doh.

Las reflexiones que se podían dar en las paredes de la habitación, y hasta en los objetos que en ella hay, pueden complicar no poco el comportamiento del sonido. Para obviarlo, Derov suspendió unas mantas formando una tienda de campaña por encima y enfrente del semicírculo. Además, sustituyó el silbato y la vela de Rayleigh por otros elementos. En un conjunto de ensayos usó un altavoz de 51 milímetros como fuente de sonido y un micrófono como detector. En otra tanda trabajó con dos osciladores piezoeléctricos de 76 milímetros; uno lo usaba como fuente y otro como detector. También, aunque brevemente, ensayó con un par de pequeños auriculares del tipo de los que se venden con algunos televisores, pero su rango de frecuencias era bastante estrecho. En todos los casos, la fuente sonora emitía una onda sinusoidal procedente de un audiooscilador con un controlador de volumen y frecuencia; la señal procedente del detector se leía en un osciloscopio.

Existe una gran variedad de equipos válidos para su empleo en estos experimentos. Lo primero que se necesita es una fuente emisora capaz de emitir frecuencias en un rango que varíe desde un kilohertz hasta 25 kilohertzs, aproximadamente. El detector no necesita tener más que un medidor de volumen como el que hay en algunos magnetofones. Se debe evitar el uso de grandes objetos en el montaje del experimento, ya que las reflexiones de las ondas sonoras sobre tales objetos pueden complicar los datos. Los transductores piezoeléctricos se venden normalmente en las tiendas de suministros de equipos electrónicos y de radio, así como los altavoces de alta frecuencia para los sistemas estéreos; el costo normal de cada uno de ellos es de unos 8 dólares para un rango de frecuencias de 7 a 25 kilohertzs.

Para cada conjunto de experimentos, Derov situaba tanto la fuente emisora como el detector a la misma altura (aproximadamente hacia la mitad de la pared). El emisor lo colocaba paralelo a la pared y el detector de igual manera, pero mirando hacia el extremo en donde estaba situado el emisor; de vez en cuando, reorientaba el detector para comprobar si el sonido se propagaba directamente a través del semicírculo. Como Rayleigh, Derov colocaba el emisor y el detector cada uno en un extremo, y, siguiendo sus pasos, encontró que el sonido era más intenso en la capa inmediata a la pared. Observó también que el sonido se eliminaba fácilmente colocando una tablilla estrecha en cualquier punto de la semicircunferencia, entre el emisor y el detector. Según parece, Rayleigh y Derov hallaron que las ondas so-

nes. El mismo Herschel escribió: "este punto fue descubierto por accidente y la persona que lo hizo se complugo durante algún tiempo en escuchar, y en llevar a sus amigos con él a oír, confesiones que concernían exclusivamente al sacerdote. Se dice que un día acudió a confesarse su esposa y tanto él como sus amigos tuvieron acceso a ciertos secretos que eran todo lo contrario de divertidos para uno de ellos".

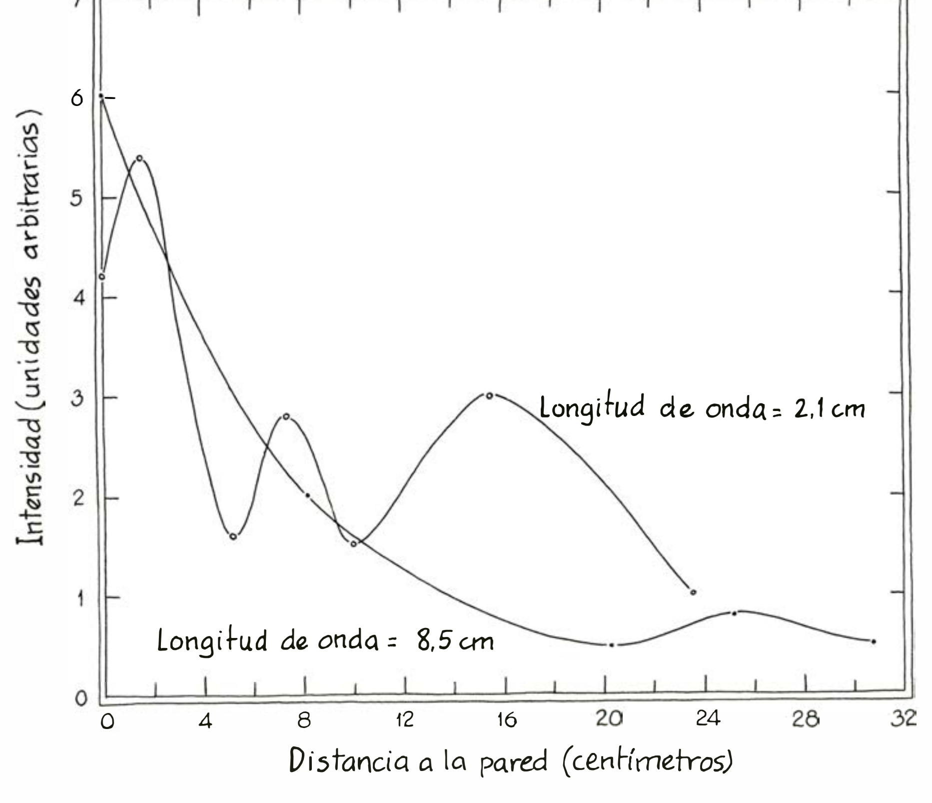
Se supone que existen "salas de los secretos" de los dos tipos generales debajo de los arcos bien pulidos de algunos puentes. En 1948 fueron analizados algunos de estos arcos separadamente por Herbert Grove Dorsey y Arthur Taber Jones para comprobar si las propiedades acústicas se debían a las ondas superficiales de Rayleigh. Por ejemplo, Dorsey investigó el arco existente bajo el puente de la Avenida Massachusetts sobre Rock Greek, en Washington. Este arco circular tiene un diámetro de 45 metros y una anchura de 22,5 metros. Ambos extremos terminan en una base de hormigón liso que produce ecos penetrantes. De acuerdo con la suposición de Rayleigh, las ondas rebotarían de un lado al otro del arco a medida que se adhiriesen a la superficie interior. Dorsey sostenía que si uno se colocaba próximo a la pared del arco aprovechando la quietud existente por la mañana, se podían oír dos ecos distintos: si la fuente sonora era un ligero susurro o la caída de un alfiler dentro de una cajita de cartón desde una altura de algo más de un centímetro. Una palmada producía de 5 a 10 ecos.

Más tarde, Jones investigó sobre un puente similar en Newton Upper Falls (Massachusetts), pero no pudo confirmar la hipótesis de Dorsey de que el sonido se propagaba a lo largo de un arco como lo hacen las ondas de superficie de Rayleigh. Jones sostenía que si estuvieran implicadas las ondas de superficie, un observador colocado en el punto central bajo el arco no oiría los ecos que se producirían cuando otra persona diera palmadas en un extremo del mismo, ya que el sonido estaría limitado a un cinturón relativamente estrecho justo debajo de la superficie del arco. Poniendo a un observador debajo del puente, Jones descubrió que los ecos eran aparentemente tan abundantes allí como en los extremos del arco. Así pues, la contribución de las ondas superficiales a los ecos múltiples quedó sin concretar.

Algunas formaciones naturales tienen propiedades similares a las de las "salas de los secretos". La Oreja de Dionisio, una gruta de una vieja cantera cercana a Siracusa en Sicilia, es una de ellas. La gruta, en forma de *S* invertida y horizontal, mide 67 metros de longitud por 22 metros de altura. La anchura de la gruta es de unos 10 metros en el suelo, pero la *S* se va estrechando hacia arriba hasta que mengua a uno o dos metros. En el extremo de la *S* opuesto a la cantera, cerca del techo de la gruta, una pequeña abertura conduce a un corto pasadizo que, a su vez, desemboca en una escalera de piedra que llega a la cima de la formación rocosa. Cuando se está de pie en el suelo de la gruta, los ecos de las palabras son tan numerosos que es difícil seguir una conversación. Sin embargo, si se escucha desde la pequeña abertura de la parte posterior de la gruta se pueden oír incluso los susurros de otras personas que estén en el suelo de la gruta. Aun cuando la gruta no sea, en su sentido estricto, una "sala de los secretos", su estructura canaliza de alguna manera el sonido desde el suelo hasta una persona que esté situada en la pequeña abertura de la parte posterior.

El origen de su nombre —Oreja de Dionisio— arranca de cuando el célebre tirano de Siracusa, Dionisio, hizo de la gruta una prisión. Según la leyenda, tenía una diseñada de tal forma que podía oír incluso los leves susurros de los encarcelados y, así frustrar cualquier intento de revuelta.

Si el lector quisiera buscar "salas de los secretos", le sugeriría que probara en cualquier sección de esfera o elipsoide en edificios y formaciones naturales, tales como cúpulas en los techos de bodegas. Se podrían investigar también las arcadas lisas (tales como las existentes bajo los puentes), los interiores de silos y otras grandes estructuras con sección circular. Una vez descubierta una "sala de los secretos", se buscan las ondas superficiales de Rayleigh por medio de un apropiado juego de experimentos, por ejemplo: colocar una barrera relativamente estrecha en la superficie curvada para bloquear cualquier onda superficial.



Resultados de Derov con dos longitudes de ondas

En septiembre de 1977 se describió en esta misma sección un experimento para la creación de arco iris de orden múltiple en el laboratorio; a lo largo de los meses siguientes recibí bastantes cartas de comunicantes que decían haber visto más de dos arcos iris naturales simultáneamente en el cielo. El arco iris primario se produce por una simple reflexión de la luz dentro de una gota de agua al caer. El poco frecuente arco iris secundario se produce por dos reflexiones internas. En el experimento se pueden ver hasta una docena de arcos iris que se forman por un número progresivamente mayor de reflexiones internas en gotas de agua suspendidas de un cable fino.

Yo siempre creía que era imposible ver arcos iris naturales de orden superior al segundo porque sus colores se hacen tan pálidos que desaparecen debido al brillo del cielo, a la luz reflejada por la superficie exterior de las gotas y a la luz transmitida a través de las gotas sin sufrir ninguna reflexión interna. Sin embargo, se ha probado pueden verse arcos iris de orden superior bajo determinadas circunstancias.

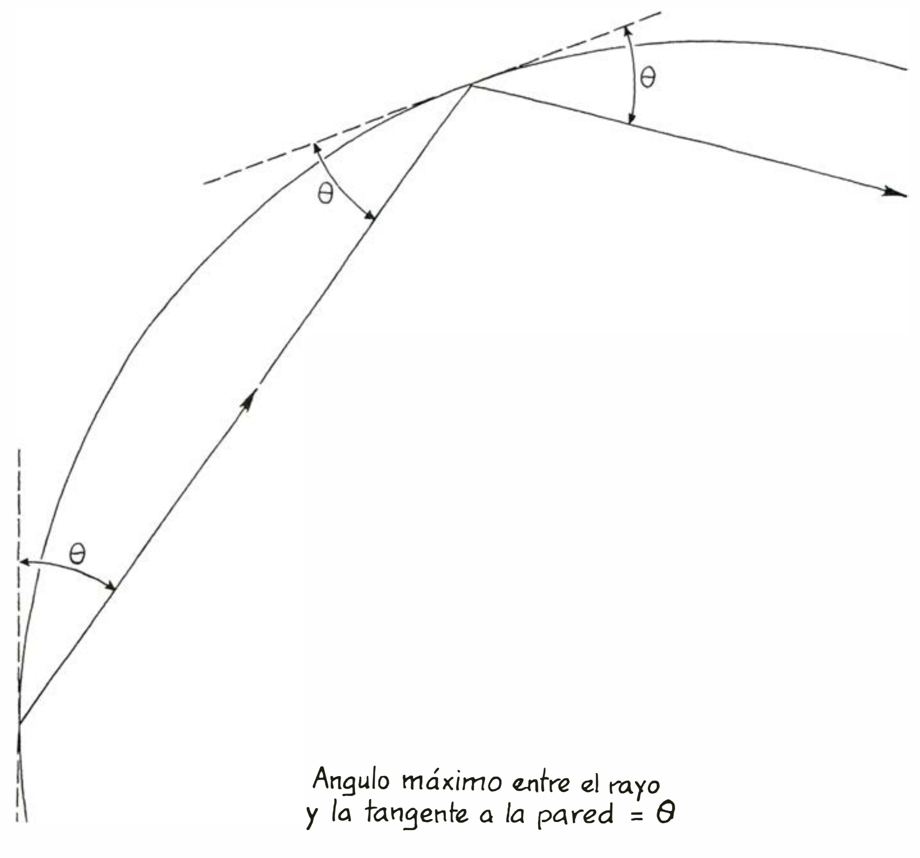
gar el semicírculo en el exterior, con la concavidad hacia arriba, en un amplio espacio abierto. También se podría construir un semicírculo mayor o usar una estructura arquitectónica existente que tenga un radio de curvatura grande. Si no se consigue una estructura mayor, habría que alcanzar frecuencias más altas, de suerte que las longitudes de onda sean mucho más cortas que el radio de curvatura.

Además de tomar datos como los de Derov, quizás interese saber cómo cambia la intensidad de las ondas de la superficie con la altura a lo largo de la pared, y cómo cambia de intensidad el modelo si la fuente emisora y el detector no se hallan a la mitad de la altura de la pared. Una característica notable de la galería de St. Paul es que la pared está ligeramente inclinada hacia dentro. Algunos investigadores han sugerido que esta inclinación ayuda a mantener las ondas de superficie. Si uno mismo construye una pared con una ligera inclinación podría confirmar esta hipótesis. Se supone que el efecto de las "salas de los secretos" también funciona con estructuras no rígidas (aunque el modelo matemático es más complicado) y por ello se podría realizar el experimento con paredes finas para hacer comparaciones.

Si quien realiza este experimento es un aficionado a la electrónica, quizá pretenda medir la velocidad de las ondas superficiales en un sistema normal. Se mueven algo más lentamente que el sonido en una línea recta; sería muy interesante comprobar la diferencia de velocidad existente entre ambos. Esto se podría medir usando un oscilador que emitiera un pulso de sonido corto (se debe hace lo más próximo que se pueda a una frecuencia única) y también pusiera en marcha un osciloscopio. Conéctese el detector al osciloscopio. El tiempo necesario para que la vibración llegue al detector comienza cuando se inicia la señal en el osciloscopio y termina cuando empieza el modelo de oscilación.

Si el aficionado tuviera la oportunidad de viajar a Londres, debería visitar la "sala de los secretos" de la catedral de St. Paul. Cuando yo la visité, aunque ya sabía lo que iba a observar, quedé maravillado de sus sorprendentes condiciones acústicas. Una amiga mía se puso en el extremo opuesto del corredor y empezó a hablar de cara a la pared. Pude oír sus susurros y conversación normal, a pesar del gran ruido que hacían los turistas tanto en el corredor como, más abajo, en los bancos. (Es mejor visitar la galería en invierno, cuando hay menos turistas.)

El efecto parecía más pronunciado cuando mi amiga y yo permanecimos próximos a la pared, pero en una conversación normal el efecto subsistía cuando uno o ambos nos movíamos hacia la barandilla existente a lo largo del corredor. Siguiendo el consejo de un experto en acústica arquitectónica, Robert S. Shankland, hablé con mi amiga usando palabras elegidas cuyo sonido produjera alta o baja frecuencia. Como ya me imaginaba, las palabras dominadas por ondas de alta frecuencia eran las más fáciles de entender.



Detalle del modelo

Mientras estaba en la galería noté que la pared no sólo estaba inclinada hacia dentro, sino que tenía además un gran borde en la parte alta. Shankland cree que estos dos hechos deben ser responsables, en gran parte, de las singulares características acústicas de la galería. Ninguna otra de las muchas galerías circulares existentes en el mundo que carecen de estas características poseen las mismas aptitudes acústicas.

Existen "salas de los secretos" del tipo de reflexión simple en muchos lugares. Algunas, tales como el viejo vestíbulo de la Cámara de Representantes del Capitolio de Washington, se han hecho famosas por su capacidad de trasmitir incluso leves murmullos. En 1851 se construyó el vestíbulo dentro de la estancia de las Estatuas y se convirtió, como indicó el experto en acústica Wallace C. Sabine, en "una de las más perfectas salas de los secretos". El techo era una sección de esfera, cuyo centro de curvatura caía casi a la altura de la cabeza. Cuando alguien se ponía en el centro y hablaba en voz baja, el sonido hacía eco y volvía al centro, concentrado por la superficie esférica. A veces, los guías colocaban a una persona en un extremo y a otra en el opuesto, a una distancia casi igual del centro, donde un murmullo de una sería reflejado y concentrado parcialmente hacia la otra. En 1901, tras un incendio, se procedió a la restauración del vestíbulo y se cambió el techo de madera pulida por otro de acero y escayola, junto con paneles suspendidos, altorrelieves y nervaduras. Como consecuencia de la abundancia de entrantes y salientes en el techo se redujo mucho la simple reflexión y concentración y el vestíbulo perdió puntos en su calificación de "sala de los secretos".

El santuario mormón en Salt Lake City constituye un ejemplo de "sala de los secretos" elipsoidal, aunque la forma interior no sea un elipsoide perfecto. Los sonidos procedentes de la mesa del lector situada en la parte frontal se transmiten por reflexión y se concentran hacia la parte de atrás, haciendo que éstos sean audibles en la parte delantera del anfiteatro posterior.

Según Herschel una reflexión similar en un interior elipsoidal fue causa de gran perplejidad en Sicilia poco después de ser construida la vieja catedral de Girgenti. Aunque los detalles del interior fueron examinados más tarde por Sabine, Herschel mantiene que uno de los focos interiores fue elegido inadvertidamente para ubicar el confesionario; por ello, cualquier persona que estuviera en el otro foco podía oír incluso las más secretas confesio-

nes. El mismo Herschel escribió: "este punto fue descubierto por accidente y la persona que lo hizo se complugo durante algún tiempo en escuchar, y en llevar a sus amigos con él a oír, confesiones que concernían exclusivamente al sacerdote. Se dice que un día acudió a confesarse su esposa y tanto él como sus amigos tuvieron acceso a ciertos secretos que eran todo lo contrario de divertidos para uno de ellos".

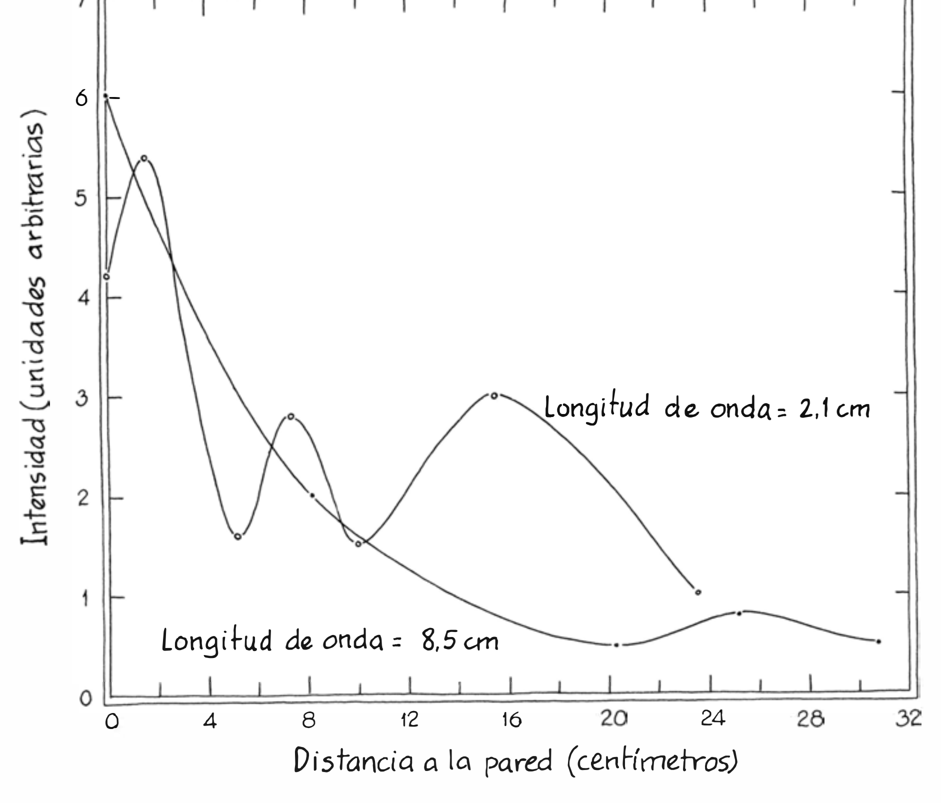
Se supone que existen "salas de los secretos" de los dos tipos generales debajo de los arcos bien pulidos de algunos puentes. En 1948 fueron analizados algunos de estos arcos separadamente por Herbert Grove Dorsey y Arthur Taber Jones para comprobar si las propiedades acústicas se debían a las ondas superficiales de Rayleigh. Por ejemplo, Dorsey investigó el arco existente bajo el puente de la Avenida Massachusetts sobre Rock Greek, en Washington. Este arco circular tiene un diámetro de 45 metros y una anchura de 22,5 metros. Ambos extremos terminan en una base de hormigón liso que produce ecos penetrantes. De acuerdo con la suposición de Rayleigh, las ondas rebotarían de un lado al otro del arco a medida que se adhiriesen a la superficie interior. Dorsey sostenía que si uno se colocaba próximo a la pared del arco aprovechando la quietud existente por la mañana, se podían oír dos ecos distintos: si la fuente sonora era un ligero susurro o la caída de un alfiler dentro de una cajita de cartón desde una altura de algo más de un centímetro. Una palmada producía de 5 a 10 ecos.

Más tarde, Jones investigó sobre un puente similar en Newton Upper Falls (Massachusetts), pero no pudo confirmar la hipótesis de Dorsey de que el sonido se propagaba a lo largo de un arco como lo hacen las ondas de superficie de Rayleigh. Jones sostenía que si estuvieran implicadas las ondas de superficie, un observador colocado en el punto central bajo el arco no oiría los ecos que se producirían cuando otra persona diera palmadas en un extremo del mismo, ya que el sonido estaría limitado a un cinturón relativamente estrecho justo debajo de la superficie del arco. Poniendo a un observador debajo del puente, Jones descubrió que los ecos eran aparentemente tan abundantes allí como en los extremos del arco. Así pues, la contribución de las ondas superficiales a los ecos múltiples quedó sin concretar.

Algunas formaciones naturales tienen propiedades similares a las de las "salas de los secretos". La Oreja de Dionisio, una gruta de una vieja cantera cercana a Siracusa en Sicilia, es una de ellas. La gruta, en forma de *S* invertida y horizontal, mide 67 metros de longitud por 22 metros de altura. La anchura de la gruta es de unos 10 metros en el suelo, pero la *S* se va estrechando hacia arriba hasta que mengua a uno o dos metros. En el extremo de la *S* opuesto a la cantera, cerca del techo de la gruta, una pequeña abertura conduce a un corto pasadizo que, a su vez, desemboca en una escalera de piedra que llega a la cima de la formación rocosa. Cuando se está de pie en el suelo de la gruta, los ecos de las palabras son tan numerosos que es difícil seguir una conversación. Sin embargo, si se escucha desde la pequeña abertura de la parte posterior de la gruta se pueden oír incluso los susurros de otras personas que estén en el suelo de la gruta. Aun cuando la gruta no sea, en su sentido estricto, una "sala de los secretos", su estructura canaliza de alguna manera el sonido desde el suelo hasta una persona que esté situada en la pequeña abertura de la parte posterior.

El origen de su nombre –Oreja de Dionisio– arranca de cuando el célebre tirano de Siracusa, Dionisio, hizo de la gruta una prisión. Según la leyenda, tenía una diseñada de tal forma que podía oír incluso los leves susurros de los encarcelados y, así frustrar cualquier intento de revuelta.

Si el lector quisiera buscar "salas de los secretos", le sugeriría que probara en cualquier sección de esfera o elipsoide en edificios y formaciones naturales, tales como cúpulas en los techos de bodegas. Se podrían investigar también las arcadas lisas (tales como las existentes bajo los puentes), los interiores de silos y otras grandes estructuras con sección circular. Una vez descubierta una "sala de los secretos", se buscan las ondas superficiales de Rayleigh por medio de un apropiado juego de experimentos, por ejemplo: colocar una barrera relativamente estrecha en la superficie curvada para bloquear cualquier onda superficial.



Resultados de Derov con dos longitudes de ondas

En septiembre de 1977 se describió en esta misma sección un experimento para la creación de arco iris de orden múltiple en el laboratorio; a lo largo de los meses siguientes recibí bastantes cartas de comunicantes que decían haber visto más de dos arcos iris naturales simultáneamente en el cielo. El arco iris primario se produce por una simple reflexión de la luz dentro de una gota de agua al caer. El poco frecuente arco iris secundario se produce por dos reflexiones internas. En el experimento se pueden ver hasta una docena de arcos iris que se forman por un número progresivamente mayor de reflexiones internas en gotas de agua suspendidas de un cable fino.

Yo siempre creía que era imposible ver arcos iris naturales de orden superior al segundo porque sus colores se hacen tan pálidos que desaparecen debido al brillo del cielo, a la luz reflejada por la superficie exterior de las gotas y a la luz transmitida a través de las gotas sin sufrir ninguna reflexión interna. Sin embargo, se ha probado pueden verse arcos iris de orden superior bajo determinadas circunstancias.

Libros

Los problemas de una sociedad cibernética, nueva tecnología pulvimetalúrgica y la química cuántica

Antonio Alabau, Enrique Julve y Manuel Fernández

ORDENADORES Y SOCIEDAD CIBERNETICA por Michael A. Arbib. Editorial AC; Madrid, 1978; 509 páginas. Con una diferencia de muy pocos meses de publicación del original, aparece la versión castellana de este libro, que como dice el autor en su prólogo "se ha escrito para explicar el impacto que los ordenadores tienen sobre nosotros hoy, y para proporcionar los conocimientos necesario para ayudar al lector a hacerse una idea de cómo se utilizarán los ordenadores en el futuro".

Ya de entrada, el título sugiere un primer comentario; el término ordenador o computador ha sido, hasta ahora, sinónimo de elevada potencia de tratamiento de información y también de *elevado precio*; como resultado de la evolución tecnológica se está modificando substancialmente esta concepción. El término ordenador continúa siendo sinónimo de "elevada potencia de cálculo" pero, sorprendentemente, empieza a serlo de precio reducido o al menos asequible, y esta es sin duda una de las claves de la segunda parte del título: "Sociedad Cibernética".

Desde que en 1948 Norbert Wiener utilizó la palabra cibernética para denominar "el estudio del control y la comunicación en el animal y en la máquina", han aparecido un gran número de definiciones que tratan de poner de manifiesto, de una forma u otra, la relación entre el comportamiento y la teoría de control. Incluso ha llegado a definirse la cibernética como "el arte de asegurar la eficacia de la acción".

Al utilizar la palabra cibernética, calificando a sociedad, se crea un nuevo término que necesita una explicación; el autor define la sociedad cibernética como "aquella en la que las personas y las máquinas interaccionan para procesar grandes cantidades de información". La definición es correcta; de ella se deduce que el ordenador constituye el ingrediente necesario para convertir en cibernética una sociedad, y lógicamente tanto más lo será ésta cuanto mayor sea la utilización del ordenador por el cuidadano. Hacia eso se tiende, tanto por la disminución de su precio, como por la simplificación de su manejo.

Pero en nuestra opinión, esta definición, siendo correcta resulta, quizá deliberadamente, inocua y, hasta cierto punto ingenua, pues nada dice sobre el objetivo de dicho "proceso de información" para el que interaccionan hombres y máquinas; no debe olvidarse la proximidad entre los conceptos de cibernética y control, y en consecuencia, lo cerca que puede estar la expresión sociedad cibernética de la expresión sociedad controlada.

El propio autor define el control como "la elección de las entradas a un sistema para hacer que el estado o la salida (en definitiva su comportamiento) cambien de un modo deseado, o al menos próximo a él"; es pues evidente que la clave del problema en la sociedad cibernética reside en quién tiene la posibilidad real de decidir sobre la elección de las "entradas" a este "sistema", sin duda complejo, y determinar de ese modo su comportamiento; en definitiva, quién ejerce esta nueva forma de poder.

Hemos observado que, en la traducción del título, se ha omitido el artículo que precedía a sociedad cibernética en su versión original, dándole un tono de indeterminación que puede contribuir a hacer pensar al lector que se trata de una sociedad distinta o al menos lejana a la suya; nada más incierto. Nuestra sociedad se está "cibernetizando" a pasos agigantados, y hay que prever sus consecuencias. Reflexione el lector sobre este tema, y relaciónelo con la mención explícita que, dentro del hasta ahora proyecto de texto de la nueva Constitución española, se hace del hecho de que "la ley limitará el uso de la información de manera que quede a salvo el respeto a la intimidad personal y familiar y al hacer de los ciudadanos", lo que no es otra cosa que reconocer a nivel constitucional que únicamente quede protegido de una de sus consecuencias, como se deduce de la lectura del libro que nos ocupa.

De los 8 capítulos que constituyen la obra, el autor dedica tres (unas 200 páginas) a la explicación del funcionamiento de un ordenador y de los elementos que en él intervienen. Partiendo prácticamente de cero, va cubriendo un amplio espectro que abarca desde las ideas elementales de programa, almacenamiento de la información, técnicas de programación, conceptos de traductores y sistemas operativos, hasta llegar a las redes de ordenadores. Esta, digamos, primera parte es suficiente e incluso tiene entidad de un texto de introducción a los ordenadores; pretende dar al lector una base sólida para la comprensión de los capítulos siguientes, que constituyen el objetivo fundamental del libro.

En los cinco capítulos restantes (300 páginas), M. A. Arbib intenta presentar cinco aspectos diferentes, y a la vez complementarios, del presente y el futuro de la sociedad cibernética: la simulación de sistemas complejos, la problemática de los bancos de datos, la inteligencia artificial, los ordenadores en la enseñanza y en la fábrica y los ordenadores, la política y la planificación.

En cada uno de los temas, el autor explica una determinada técnica, muestra cómo se utiliza dicha técnica para la resolución de una clase de problemas y, por último, comenta posibles pros y contras derivados de lo tratado.

En el capítulo dedicado a la simulación, se introducen las ideas básicas de la realimentación y la teoría de control, se explica su aplicación a un problema con dos varibles –equilibrio entre una población de lobos y conejos- problema aparentemente trivial, pero del que se obtiene un sorprendente rendimiento, y por último, se especula con lo que podría ser una simulación del mundo en lo que se refiere a la población y a su control.

Al analizar la problemática de las bases de datos, se presentan someramente las características de un sistema de gestión de base de datos, se explica su influencia sobre el manejo de fondos bancarios comentando los llamados delitos "informáticos" y, por último, dedica una buena parte del capítulo al tema de la relación entre los bancos de datos y la privacidad de la información. Presenta y comenta la ley que en 1974 se promulgó en los Estados Unidos sobre la intimidad y que, entre otras cosas, regula la utilización de los sistemas de información y el acceso del ciudadano a las bases de datos que almacenan información que le concierne.

En el capítulo dedicado a la inteligencia artificial, se comenta la estructura del cerebro, se presentan distintas facultades de que puede ser dotado actualmente un robot artificial, y, por último, se trata, con una frialdad asombrosa, el tema de la utilización del ordenador en el campo de ba-

talla. Cuando describe el papel del ordenador en la enseñanza y en la fábrica, trata ampliamente el tema de la enseñanza asistida por ordenador y del control del flujo de la información en la sociedad (sorprende una vez más, el tono ingenuo con que se explica el manejo de la información y el uso que de ella hacen los estados totalitarios); por último se habla de la automatización industrial, de las consecuencias laborales, así como de las acciones sindicales frente a la automatización. También en este último tema peca el autor de ingenuo por la presentación simplista que hace de los problemas del mundo laboral, dando la impresión de que realmente quedan muy lejos, tanto de su ambiente habitual, como del de las personas a las que se dirige.

Por último, el capítulo consagrado al ordenadores en la planificación y del ordenador y la democracia, finalizando con ordenadores en la planificación, del ordenador y la democracia, finalizando con una apoteosis de política-ficción cargada de la misma dosis de infantilismo de que el autor hace gala en anteriores captítulos.

En definitiva, en el aspecto técnico y científico, el libro es serio dentro de su objetivo de divulgación, y el autor demuestra conocer muy bien el tema; pero en el aspecto sociológico, con independencia de la coincidencia, o no, con las ideas del autor, nos atrevemos a decir que el libro es flojo, como hemos pretendido poner de manifiesto.

Respecto a su destinatario, dice Arbib que "se adapta especialmente a un curso de ordenadores y sociedad a nivel de primer ciclo universitario". Personalmente opino que en el aspecto técnico podría acoplarse a un curso de tipo de los que el autor se refiere en el caso de que existiera alguno en nuestras universidades. Ahora bien, en el aspecto sociológico el nivel es muy inferior, por la forma superficial con que se tratan algunos temas, incluso por el contenido de los ejercicios que se proponen al alumno, de los que nos permitimos incluir un ejemplo verdaderamente extremo: "Enumérese consecuencias positivas y negativas de la eliminación de todas las especies de insectos" (nº 3, pág. 223).

En lo referente a la estructura, el libro está bien hecho y sigue una buena técnica pedagógica; cada capítulo está subdividido en secciones, y en todos ellos se conserva la misma pauta: introducción muy breve; texto propiamente dicho, escrito en un tono ameno; escueto resumen de las ideas fundamentales aparecidas en el texto; lista de lecturas recomendadas (desgraciadamente inaccesibles para el lector medio, por tratarse, casi en su totalidad, de revistas y libros norteamericanos) y ejercicios pertinentes.

Creemos que el libro sería mucho más aprovechable si se hiciera el esfuerzo de adaptar su contenido, y evidentemente las referencias de lectura, y quizá parte de los ejercicios, a la situación real de nuestro país; de esa forma podría servir perfectamente como texto para un curso, ya no en la universidad, sino en el bachillerato que es, en nuestra opinión, donde deberían comenzar a impartirse estas enseñanzas en nuestra sociedad cibernética.

De M. A. Arbib, profesor en el Departamento de Ciencias de los Computadores y de la Información en la Universidad de Massachusetts, existe otro libro en castellano que, con el título de "Cerebros, Máquinas y Matemáticas", fue publicado por Alianza Universidad, y que apareció en 1976, esta vez, nada menos que 12 años después de que fuera escrito, en el que el lector podrá encontrar un buen resumen de las teorías que se engloban bajo las denominaciones de cibernética e inteligencia artificial, y que constituye un buen complemento del texto del que nos hemos ocupado (A. A.).

INTRODUCCION A LA PULVIMETALURGIA por Pedro Molera. Ediciones Bellaterra S.A., Barcelona 1977. Un volumen de 16 × 22 cm., encuadernado en tela, con 209 páginas y numerosas figuras y tablas. La pulvimetalurgia es un procedición que no cabe silenciar. Estas limitaciones se refieren en primer lugar, a la forma el moldeo. Abarca todas las etapas comprendidas entre la obtención de los polvos metálicos y la obtención de la pieza enteramente acabada; es decir, las etapas de: producción de polvos, mezcla, aglomeraciones, sinterización y acabados. Por eliminar etapas en los procesos de fabricación y ahorrar pérdidas de materia prima, esta técnica metalúrgica va ganando día a día nuevos adeptos, especialmente en aquellos casos en que sólo ella es capaz de resolver problemas concretos, como pueden ser la obtención de materiales imposibles de fabricar con las técnicas metalúrgicas convencionales (caso de los cojinetes autolubricantes, los filtros metálicos, el metal duro o las ferritas).

Metalúrgicamente hablando, parece ser que la humanidad está avanzando de una manera progresiva e irreversible hacia un porvenir nada halagüeño. Tomando como modelo los países tecnológicamente desarrollados, se evidencian dos hechos contradictorios, que se producen a la vez: por un lado, el consumo creciente de metales y, por otro, el empobrecimiento gradual de las menas metálicas. Al solucionar, en parte, la pulvimetalurgia, la escasez de materia prima que de la problemática citada se deriva (al eliminar la chatarra de proceso de transformación que se produce con los métodos convencionales de hechurado), queda convertida esta técnica en una adelantada de su época y en una indudable promesa para el futuro.

La pulvimetalurgia presenta, sin embargo, algunas limitaciones en su aplicación que no cabe silenciar. Estas limitaciones se refieren en primer lugar, a la forma de la pieza a elaborar, ya que debe reunir unas condiciones en su diseño que permitan su fácil extracción de la matriz; en segundo lugar, a las dimensiones de la pieza a elaborar, limitadas a las características de las actuales prensas de tipo normal, las cuales no suelen sobrepasar las 500 Tm; en tercer lugar, a las características mecánicas a elaborar, no cabiendo pedirle las mismas cualidades de alargamiento y resistencia exigibles a las obtenidas por los métodos tradicionales y, en cuarto lugar, al costo del proceso, ya que las matrices utilizadas en pulvimetalurgia, al ser de acero aleado o de carburo de tungsteno, con acabados difíciles de conseguir, suelen resultar relativamente caras, incidiendo en bastante cuantía en el precio de las piezas obtenidas por este proceso.

A presentarnos toda la problemática apuntada, al propio tiempo que a estudiar la pulvimetalurgia bajo el doble enfoque científico y técnico práctico, aspira la obra objeto de este comentario, aspiración plenamente conseguida. Su autor, relacionado con la metalurgia tanto teórica como práctica a lo largo de sus años de vinculación con el Instituto Metalúrgico "Emilio Jimeno" y con el Departamento de Metalurgia de la Universidad de Barcelona, es un destacado especialista en la nueva técnica de la pulvimetalurgia, lo que da una garantía de la seriedad y del rigor de la obra.

El libro consta de trece capítulos y un apéndice, precedidos por un prólogo a cargo del Dr. Calvo, catedrático de Metalurgia de la Universidad Complutense y autoridad reconocida en el campo de la metalurgia teórica. En el capítulo primero, de carácter introductorio, se exponen el concepto de la pulvimetalurgia, su desarrollo histórico, su campo de aplicación actual y su futuro, En el capítulo segundo, y dada la indudable influencia que las características de los polvos poseen en la calidad del producto y en la economía del proceso de fabricación, se ofrece un estudio completo sobre las propiedades de los polvos –haciendo referencia a la forma, composición, tamaño y distribución, porosidad, microestructura, densidad, superficie específica, velocidad de derrame, comprensibilidad, consolidabilidad, plasticidad y acritud–; aporta la normativa vigente en pulvimetalurgia y los métodos de obtención de los polvos (reducción, atomización, electrolítico, tritu-

ración, pirolisis, corrosión, condensación, amalgamación, liofiliación y precipitación.

Una vez obtenidos los polvos, y antes de compactarlos, se requiere un tratamiento adecuado, tema del capítulo tercero. Allí, se describen las operaciones de purificación, de recubrimiento, de densificado, de mezclado y de recogido. El capítulo cuarto está dedicado a la etapa de compactación de los polvos, fase en que se da forma y consistencia a la masa de polvo, con objeto de que pueda manipularse en el intervalo que va desde su expulsión de la matriz hasta su paso por el horno de sinterización. Comienza el capítulo con una referencia al concepto de compactación, para seguir con la descripción de las diferentes fases de la misma –llenado, comprensión y expulsión– y finalizar con los métodos de compactación, a saber: compactación cíclica, compactación por vibración, compactación explosiva, compactación caliente, "spark sintering", compactación por presión isostática y compactación por laminación.

El capítulo quinto se consagra a la etapa que sigue a la compactación, es decir, la sinterización. La sinterización consiste en calentar la masa de polvo compactada a temperatura inferior a su punto de fusión durante un período de tiempo suficiente para que las partículas se suelden y adquieran resistencia metálica. Después de dar unas nociones generales de la sinterización, se repasan las variables de dicha etapa, los tipos de hornos que se suelen utilizar (atendiendo a la clase de calentamiento y a la masa gaseosa encerrada dentro del horno) y, finalmente, las distintas teorías que existen sobre el fenómeno de la sinterización.

Los acabados, que abarcan cada una de las operaciones que tienen por finalidad el mejorar las características mecánicas, físicas y químicas del sinterizado, al propio tiempo que el aspecto estético y el control dimensional, constituyen el tema del capítulo siguiente. Se estudia, pues, el calibrado, el mecanizado, los tratamientos térmicos, la corrosión, pavonado, los recubrimientos electrolíticos, recubrimiento mecánico y la soldadura.

Una vez descritas cada una de las etapas del procedimiento pulvimetalúrgico, el autor acomete el estudio de las técnicas especiales aplicadas con objeto de reducir la porosidad residual de las piezas obtenidas por este procedimiento, citando la técnica denominada "sinter forging", o forja de preformas sinterizadas, y la técnica de la infiltración, objetivo del capítulo séptimo. La primera de estas técnicas ha hallado especial aplicación en la producción a escala industrial de piezas diversas de automóviles (piñones de diferencial y de caja y ejes de transmisión, principalmente). La segunda técnica se ha utilizado preferencialmente en la fabricación de pseudo-aleaciones (de cobre-tungsteno y plata-tungsteno) y para mejorar propiedades mecánicas de metales, como en el caso del acero en el que se infiltra cobre.

En los siguientes capítulos (del octavo al décimosegundo), basándose en las características especialmente favorables del procedimiento pulvimetalúrgico, se detalla la aplicación del mismo a la elaboración de piezas estructurales, a la elaboración de aceros aleados y de aleaciones porosas, a la elaboración de metales oxidados internamente y, por último, a la elaboración de materiales con propiedades físicas especiales (eléctricamente conductores o con propiedades magnéticas). Así, en el apartado destinado a las piezas estructurales, se aplica la pulvimetalurgia a la obtención de aceros al carbono, de aleaciones de hierro-cobre y de la aleación especial hierro-cobre-carbono. En el siguiente, sobre los aceros aleados, se aplica la técnica a los aceros al niquel, a los aceros rápidos (que portan como elementos aleantes el tungsteno, cromo, vanadio, molibdeno y cobalto), a los aceros inoxidables (que incluyen cromo y niquel en la proporción 18Cr/8Ni, generalmente) y a los aceros "maraging" (aleaciones de muy bajo contenido en carbono que portan además niquel y molibdeno). Al estudiar las aleaciones porosas, se aplica la pulvimetalurgia a la elaboración de bronces autolibricantes, de niquel poroso, de implantes, de aleaciones antifricción y de filtros metálicos sinterizados. En el capítulo undécimo, destinado a los metales refractarios, a materiales pesados y a metales internamente oxidados, se aplica el procedimiento pulvimetalúrgico a la fabricación de componentes de tungsteno, de molibdeno y de tántalo, a la elaboración de componentes a partir de unos materiales de elevada densidad (pseudoaleaciones de tungsteno-niquel-cobre) utilizadas para la absorción de los rayos gamma y como masa inerte en sistemas en rotación y a la elaboración de ciertos metales endurecidos por oxidación interna (como pueden ser: niquel-óxido de torio, cobre-óxido de berilio o niquel-óxido de aluminio). Respecto a los materiales que gozan, de propiedades conductoras o magnéticas, la pulvimetalurgia se aplica a aleaciones conductoras y resistentes, tales como cobre-tungsteno, plata-tungsteno y plata-molibdeno y a aleaciones con propiedades magnéticas, tales como los imanes "alnife" (aluminio-niquel-hierro) y "alnico" (aluminio-niquel-cobalto-cobre-hierro).

En el capítulo subsiguiente se hace una especial referencia a la fabricación, mediante este nuevo procedimiento de conformación metálica, del llamado metal duro (carburos sinterizados), que están compuestos habitualmente por tungsteno-hierro, tungsteno-cobalto o por tungsteno-niquel. Después de estudiar los correspondientes diagramas de equilibrio y los distintos métodos de fabricación de carburos, el autor cita las propiedades mecánicas, físicas y químicas de los carburos y los diferentes factores que influyen en esas propiedades y los ensayos no destructivos y aplicaciones de este metal duro.

El autor da cuenta en el Apéndice los posibles mecanismos de la etapa de la sinterización, resumiento la demostración teórica de los mismos y escogiendo con ese fin como modelo una esfera metálica apoyada en una base plana de la misma naturaleza. Al final de cada capítulo y al agregar la correspondiente bibliografía, que enriquece el contenido de la obra, la primera en lengua castellana que trata de este interesante procedimiento metalúrgico. (E.J.)

QUIMICA CUANTICA, por Ira N. Levine. Editorial A.C. Madrid 1977. Traducido de la segunda edición estadounidense (Allyn and Bacon 1974) por A. Fuster Ortigosa y A. Requena Rodríguez. Este libro, dirigido fundamentalmente a estudiantes de los últimos años de licenciatura en ciencias químicas, viene a añadirse con gran oportunidad a los escasos textos generales de química cuántica que pueden leerse en español. En efecto, a pesar de la gran abundancia de bibliografía extranjera, hasta el momento sólo había dos libros de este tipo traducidos a nuestro idioma: *Química Cuántica Básica*, de L.F. Phillips, y *Teoría Cuántica de Atomos, Moléculas y Fotones*, de J. Avery (ambos textos editados por Alhambra en su conocida Colección Exedra).

La química cuántica constituye una de las ramas más modernas de la química física y trata de la aplicación de la mecánica cuántica al estudio de diversos problemas de interés químico, como los planteados en estructura atómico molecular, en espectroscopía, o en reactividad. En parte por lo reciente de su aparición y también por el rápido desarrollo de la tecnología de ordenadores con los que la química cuántica obtiene algunos de sus resultados -quizá los más espectaculares aunque, probablemente, no sean los más importantes-, la materia del libro que comentamos ha sufrido en los últimos años un crecimiento casi explosivo. A pesar de ello el autor ha logrado "rejuvenecer" completamente su libro respecto a la primera edición mediante una serie de cambios bien elegidos, sobre todo en lo que se refiere al tratamiento de los métodos de cálculo (se incluye el estudio de métodos semiempíricos como CNDO, INDO y

MINDO, y la aplicación de bases gaussianas en el perfeccionamiento técnico de los métodos *ab initio*). Asimismo, ha sido importante la separación del material que contenía la primera edición en dos partes, publicando en la segunda edición de la obra *química cuántica* sólo una de ellas, la que se refiere a las bases y métodos de esta disciplina, sin insistir demasiado en sus aplicaciones, aunque sin olvidarlas. La otra parte, que se refería casi exclusivamente a las aplicaciones de la química cuántica en la espectroscopía, se publicó independientemente, y con bastantes cambios, bajo el nombre de *Molecular Spectroscopy* (Wiley-Interscience 1975), y no ha sido traducida.

El primer problema que se presenta al proyectar un texto de química cuántica es elegir el punto en que se debe comenzar el estudio de esta materia. Los dos precedentes de la traducción que comentamos son buenos ejemplos de las dos posiciones extremas que cabe adoptar frente a esta elección. En el libro de Phillips se arranca directamente de los postulados de la mecánica cuántica, sin intentar una mentalización previa del lector en lo referente a las características del comportamiento cuántico de los electrones. En el libro de Avery, en cambio, el comienzo se remonta a los precedentes clásicos de la mecánica cuántica, lo que, si bien tiene sus ventajas, posee el grave inconveniente de restar mucho espacio al tratamiento de temas propiamente químicocuanticos.

En el libro de Levine se ha escogido un punto de partida intermedio entre los de los dos textos antes citados, consistente en empezar por una discusión no axiomática de la ecuación de Schrodinger y sus aplicaciones elementales. Esta especie de introducción constituye los dos primeros capítulos del libro. El tercer capítulo, que el autor titula "Operadores", incluye (sin resaltar todavía su carácter axiomático) los postulados de la mecánica cuántica que constituían el primer capítulo del libro de Phillips. El capítulo 4 trata, de una forma tradicional, sobre el oscilador armónico monodimensional, es decir, resolviendo la ecuación diferencial correspondiente por técnicas analíticas; en el capítulo 5 afronta el momento angular de una partícula, tratándolo tanto desde el punto de vista algebraico (operadores de ascenso y descenso basados en las propiedades de conmutación de $\hat{L}_x$, $\hat{L}_y$ y $\hat{L}_z$). El capítulo 6 estudia a los potenciales centrales, que examina haciendo uso de las propiedades del momento angular deducidas en el capítulo anterior, y el átomo de hidrógeno, que analiza con claridad exhaustiva. El capítulo 7, "Teoremas de la Mecánica Cuántica", se refiere fundamentalmente a la teoría de operadores hermíticos e incluye la introducción de las funciones delta de Dirac como funciones propias del operador coordenada y, al final, una enumeración comentada de los postulados de la mecánica cuántica que, en su mayor parte, ya habían sido introducidos de manera informal en el capítulo 3.

Los capítulos 8 y 9 estudian los métodos variacionales (8) y de perturbación (9); en cierto modo, son los últimos que podrían formar parte de un libro de mecánica cuántica ordinario. En efecto, el tratamiento del capítulo décimo, "Spin electrónico y principio de Pauli", es esencialmente químico sin profundizar apenas en el tema de la naturaleza del spin, o en las razones por las que los sistemas polielectrónicos "prefieren" funciones de onda completamente antisimétricas. Precisamente el enfoque de este tema es casi contrario en el libro de Avery, en el que se cuida más el análisis del origen del spin y sus consecuencias físicas, a cambio de dedicar muy poco espacio al estudio de sus consecuencias químicas, por lo que respecto al spin los textos de Avery y de Levine pueden considerarse complementarios.

Los capítulos 11, 13 y 15 (último) contienen, en cierto modo, las aplicaciones más importantes de la teoría desarrollada en los capítulos precedentes. El 11 se refiere a los átomos polielectrónicos, el 13 a las moléculas diatómicas y el 15 a las poliatómicas. El capítulo 12, "Simetría molecular", trata solamente de los aspectos más elementales de la teoría de grupos (en cambio el capítulo que Levine dedica a estos temas en su libro de espectroscopía molecular es muy completo). El capítulo 14 trata de los teoremas del virial y de Hellman-Feynman. Se incluye finalmente una bibliografía muy bien escogida y un gran número de problemas con su solución final, pero se echa en falta una explicación detallada de, al menos, unos cuantos de estos problemas.

En resumen, tratamos con un buen libro de introducción a la química cuántica, cuya principal virtud, sin menospreciar otras, es su claridad expositiva que llega al extremo de incluir la discusión detallada de los tópicos matemáticos que van apareciendo, como la resolución de las ecuaciones diferenciales. Precisamente es de agradecer que no se haya recurrido al método usual (y no demasiado útil) de reunir todos los "complementos matemáticos" en un apéndice o un capítulo aparte. El único defecto importante es, en nuestra opinión, la falta de algunos capítulos dedicados a los polímeros, los cristales, los compuestos de coordinación y a la espectroscopía (esta última es, como dijimos, objeto de un libro aparte del mismo autor). (M.F.N.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

TRANSMISION DIGITAL DE SEÑALES

INFORMATION AND COMUNICATION THEORY. A.M. Rosie. Van Nostrand Reinhold Company Limited.

TECNICAS NUMERICAS. Publicación Especial. Edición primera, noviembre 1975, Centro de Investigación y Estudios (C.T.N.E.).

LIBRO NARANJA, TOMO III-2, C.C.I.T.T. Sexta Asamblea Plenaria, Ginebra, 27 de septiembre-8 de octubre de 1976. U.I.T.

REDES DIGITALES Y TRANSMISION DE DATOS. Pedro Pastor Lozano. Universidad Politécnica de Madrid, E.T.S.I.T. en prensa, 1978.

FUNCION CEREBRAL Y FLUJO SANGUINEO

CORRELATION BETWEEN REGIONAL CEREBRAL BLOOD FLOW AND OXIDATIVE METABOLISM. Marcus E. Raichle, Robert L. Grubb, H. Gado Mokhtar, John O. Eichling y Michel M. Ter-Pogossian en *Archives of Neurology,* vol. 33, págs. 523-526; agosto 1976.

REGIONAL CEREBRAL BLOOD FLOW IN FOCAL CORTICAL EPILEPSY. Kjeld Hougaard. Tadato Oikawa, Edda Sveinsdottir, Erik Skinhøj. David H. Ingvar y Niels A. Lassen en *Archives of Neurology,* vol. 33, págs. 527-535: agosto, 1976.

CEREBRAL FUNCTION, METABOLISM AND CIRCULATION. Dirigido por David H. Ingvar y Niels A. Lassen en *Acta Neurologica Scandinavica Supplementum 64,* vol. 56; 1977.

LA PARTICULA HYPSILON

OBSERVATION OF A DIMUON RESONANCE AT 9.5 GEV IN 400-GEV PROTON-NUCLEUS COLLISIONS. S. W. Herb et al. en *Physical Review Letters,* vol. 39, n.º 5, págs. 522-525; 1 de agosto de 1977.

PARTICULAS FUNDAMENTALES CON ENCANTO. Roy F. Schwitters en *Investigación y Ciencia,* n.º 15, págs. 32-48; diciembre, 1977.

OBSERVATION OF STRUCTURE IN THE Y REGION. W. R. Innes et al. en *Physical Review Letters,* vol. 39, n.º 20, págs. 1240-1242; 14 de noviembre de 1977.

EL ULTIMO TEOREMA DE FERMAT

HISTORY OF THE THEORY OF NUMBERS. Leonard E. Dickson. Carnegie Institution of Washington, 1919.

THE FERMAT AND HESSIAN POINTS OF A TRIANGLE. H. E. Fettis en *The American Mathematical Monthly,* vol. 53, n.º 2, págs. 74-78; febrero, 1946.

A SUPPLEMENTARY NOTE TO A 1946 ARTICLE ON FERMAT'S LAST THEOREM. H. S. Vandiver en *The American Mathematical Monthly,* vol. 60, n.º 3, págs. 164-166; marzo, 1953.

¿DESENCADENO UNA SUPERNOVA LA FORMACION DEL SISTEMA SOLAR?

A COMPONENT OF PRIMITIVE NUCLEAR COMPOSITION IN CARBONACEOUS METEORITES. R. N. Clayton, L. Grossman y T. K. Mayeda en *Science,* vol. 182, págs. 485-488; 2 de noviembre de 1973.

ALUMINUM-26 IN THE EARLY SOLAR SYSTEM: FOSSIL OR FUEL? Typhoon Lee, D. A. Papanastassiou y G. J. Wasserburg en *Astrophysical Journal,* vol. 211, n.º 2, parte 2, págs. L107-L110; 15 de enero de 1977.

SUPERNOVAE, GRAINS AND FORMATION OF THE SOLAR SYSTEM. James M. Lattimer, David N. Schramm y Lawrence Grossman en *Nature,* vol. 269, págs. 116-118; 8 de septiembre de 1977.

LOS DIBUJOS PREHISTORICOS EN EL SUELO DEL PERU

ANCIENT DRAWINGS ON THE DESERT OF PERU. Paul Kosok y Maria Reiche en *Archaeology,* vol. 2, n.º 4, págs. 206-215; diciembre, 1949.

LIFE, LAND AND WATER IN ANCIENT PERU. Paul Kosok. Long Island University Press, 1965.

GIANT GROUND FIGURES OF THE PREHISTORIC DESERTS. Emma Lou Davis y Sylvia Winslow en *Proceedings of the American Philosophical Society,* vol. 109, págs. 8-21; 1965.

WAS HUARI A STATE? William H. Isbell y Katharina J. Schreiber en *American Antiquity,* vol. 43, n.º 3, págs. 372-389; 1978.

LA FORMACION DE MODELOS EN EL DESARROLLO BIOLOGICO

POSITIONAL INFORMATION AND PATTERN FORMATION. Lewis Wolpert en *Current Topics in Developmental Biology,* vol. 6, págs. 183-224; 1971.

SYMPOSIUM ON CELL PATTERNING. Ciba Foundation Symposium 29. Associated Scientific Publishers, 1975.

PATTERN FORMATION IN EPIMORPHIC FIELDS. Vernon French, Peter J. Bryant y Susan V. Bryant en *Science,* vol. 193, págs. 969-980; 10 de septiembre de 1976.

UNA MIGRACION EN MASA DE AVES TERRESTRES SOBRE EL OCEANO

BIRD MIGRATION. Donald R. Griffin. Natural History Press, 1964.

RADAR ORNITHOLOGY. Eric Eastwood. Methuen and Co., 1967.

AUTUMNAL BIRD MIGRATION OVER THE NORTHWESTERN ATLANTIC OCEAN. Timothy C. Williams, Janet M. Williams, Leonard C. Ireland y John M. Teal en *American Birds,* vol. 31, n.º 3, págs. 251-267; mayo, 1977.

JUEGOS MATEMATICOS

EXCURSIONS INTO MATHEMATICS. Anatole Beck, Michael N. Bleicher y Donald W. Crowe. Worth Publishers, 1969.

MATHEMATICS IN THE TIME OF THE PHARAOHS. Richard J. Gillings. The MIT Press. 1972.

TALLER Y LABORATORIO

THE ECHOES AT ECHO BRIDGE. Arthur Taber Jones en *Journal of the Acoustical Society of America,* vol. 20, n.º 5, págs. 706-707; septiembre, 1948.

THE PROBLEM OF THE WHISPERING GALLERY. John William Strutt, Baron Rayleigh, en *Scientific Papers,* vol. V, 1902-1910. Dover Publications, 1964.

WHISPERING GALLERIES. Tom Beer en *Contemporary Physics.* vol. 16, n.º 3, págs. 257-262; mayo, 1975.

ERRATUM

En la página 1 del número 23, agosto de 1978, donde dice "elementos de luz", debe decir "elementos ligeros", en el artículo de Cerny y Poskanzer.

Indice anual

Este índice contiene los nombres de los autores y artículos aparecidos en Investigación y Ciencia *a partir de enero de 1978. Se incorpora asimismo la relación alfabética de las secciones* Juegos matemáticos *y* Taller y laboratorio.

AUTORES

ARTICULOS

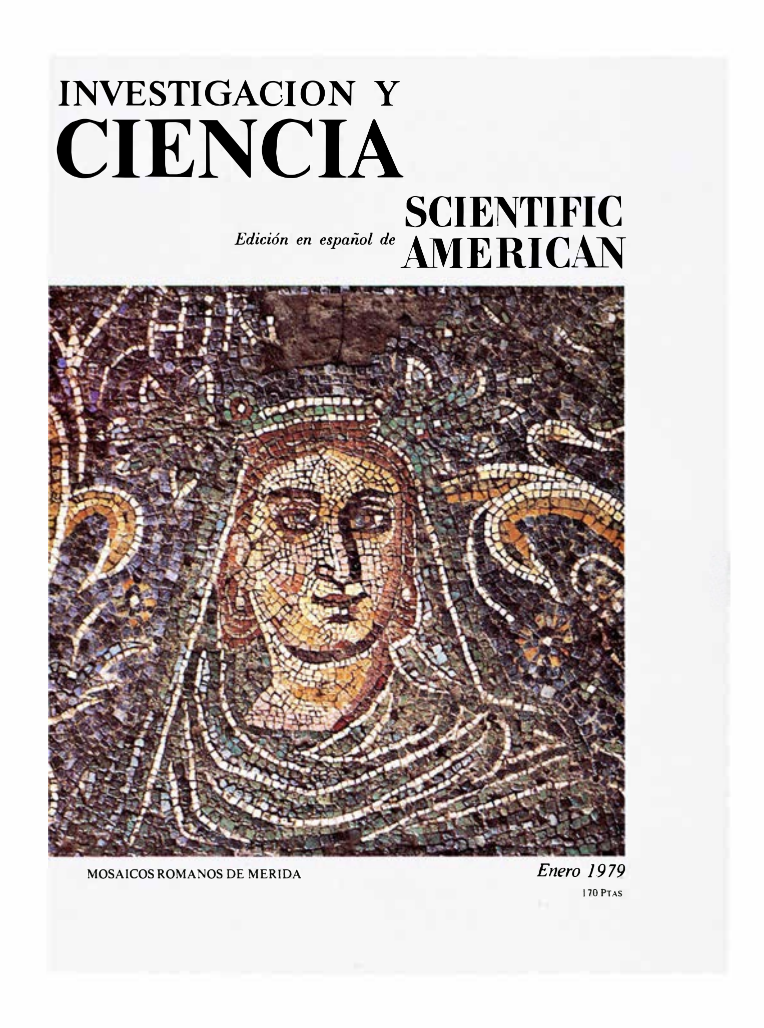
JUEGOS MATEMATICOS

Autor: Martin Gardner

TALLER Y LABORATORIO

Autor: Jearl Walker

Seguiremos explorando los campos del conocimiento



MOSAICOS ROMANOS DE MERIDA, por Antonio Blanco
La dificultad de arrancar un mosaico sin destruirlo y el poco valor de sus materiales cuando éstos se desgranan han favorecido la conservación in situ de una cantidad de mosaicos muy superior a la de otros restos de la antigüedad romana en España. Mérida, capital de la Lusitania romana, posee una riqueza excepcional en obras de arte de este género.

CUMULOS RICOS DE GALAXIAS, por Paul Gorenstein y Wallace Tucker
Alrededor de un 10 por ciento de la totalidad de las galaxias pertenece a cúmulos ricos, sistemas compuestos de miles de miembros envueltos en gas caliente. Tales cúmulos son remolinos gravitatorios con galaxias gigantes cercanas a su centro.

EL ENSAMBLAJE DE UN VIRUS, por P. Jonathan G. Butler y Aaron Klug
El virus del mosaico del tabaco, objeto del presente artículo, está constituido por una cadena de ácido nucleico, contenido en una formación tubular proteica específica. Ambos componentes se unen de forma espontánea, de acuerdo con un mecanismo complejo.

ENERGIA DE FUSION CON HACES DE PARTICULAS, por Gerold Yonas
En una aproximación a la fusión controlada se hace implosionar un gránulo de combustible mediante una fuente externa de energía. El proporcionar la energía con haces intensos de partículas está siendo estudiado en los Estados Unidos y en la Unión Soviética.

LOS MECANISMOS DEL MECANIZADO POR ABRASION, por Leonard E. Samuels
Muchos mecanizados no se efectúan por corte, sino por abrasión. Con la finalidad de mejorar la efectividad de los procesos por abrasión, se muestra la forma de trabajar de los granos abrasivos.

EL SISTEMA DE RECOMPENSA DEL CEREBRO, por Aryeh Routtenberg
Hace dos décadas se descubrió que el cerebro tiene "centros de placer". Hoy se cree que estos centros pertenecen a un sistema de vías que, al parecer, inciden en el aprendizaje y la memoria.

LA OPTICA DE RAYOS X DE LONGITUD DE ONDA LARGA, por Eberhard Spiller y Ralph Feder
Los rayos X empleados en cristalografía y radiografías médicas son "duros" o de longitud de onda corta. En la actualidad se investiga la utilización de rayos X "blandos", de longitud de onda larga, en microscopía, astronomía y microelectrónica.

EL SISTEMA DE CONTROL DE VUELO DE LOS HERMANOS WRIGHT, por Frederick J. Hooven
Los aeroplanos sin motor de los hermanos Wright tenían un elevador en el morro, no en la cola. Aunque, el sistema parece inestable, los hermanos Wright tenían sus razones para utilizarlo.

INVESTIGACION Y
CIENCIA